

(19) World Intellectual Property Organization  
International Bureau



(43) International Publication Date  
15 August 2002 (15.08.2002)

PCT

(10) International Publication Number  
**WO 02/062379 A2**

- (51) International Patent Classification<sup>7</sup>: **A61K 39/08**
- (21) International Application Number: **PCT/IE02/00017**
- (22) International Filing Date: 11 February 2002 (11.02.2002)
- (25) Filing Language: English
- (26) Publication Language: English
- (30) Priority Data:  
2001/0137 9 February 2001 (09.02.2001) IE
- (71) Applicant (*for all designated States except US*): **THE PROVOST, FELLOWS AND SCHOLARS OF THE COLLEGE OF THE HOLY AND UNIDIVIDED TRINITY OF QUEEN ELIZABETH** [IE/IE]; Near Dublin, College Green, Dublin 2 (IE).
- (72) Inventors; and
- (75) Inventors/Applicants (*for US only*): **DOYLE, Rachael** [IE/IE]; 19 Deerpark Avenue, Castleknock, Dublin 15 (IE). **KELLEHER, Dermot** [IE/IE]; 30 Royal Terrace West, Dun Laoghaire, County Dublin (IE). **WINDLE, Henry, J.** [IE/IE]; 15 Cherryfield Avenue Upper, Ranelagh, Dublin 6 (IE). **WALSH, James, Bernard** [IE/IE]; 3 Ardlui Park, Blackrock, County Dublin (IE). **DEIRDRE, Ni, Eidhin** [IE/IE]; 15 Watkins Buildings, The Coombe, Dublin 8 (IE).
- (74) Agent: **O'BRIEN JOHN A AND WELDON, Michael J**; c/o John A. O'Brien & Associates, Third Floor, Duncairn House, 14 Carysfort Avenue, Blackrock, County Dublin (IE).
- (81) Designated States (*national*): AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EC, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NO, NZ, OM, PH, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VN, YU, ZA, ZM, ZW.
- (84) Designated States (*regional*): ARIPO patent (GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE, TR), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).
- Published:**  
— *without international search report and to be republished upon receipt of that report*
- For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.*

(54) Title: CLOSTRIDIUM DIFFICILE VACCINE

(57) Abstract: A vaccine for the treatment or prophylaxis of *C. difficile* associated disease comprises a *C. difficile* gene or a *C. difficile* peptide/polypeptide or a derivative or fragment or mutant or variant thereof which is immunogenic in humans. The gene encodes a *C. difficile* surface layer protein, SlpA or variant or homologue thereof. The peptide/polypeptide is a *C. difficile* surface layer protein, SlpA or variant or homologue thereof. The vaccine may comprise a chimeric nucleic acid sequence.



**WO 02/062379 A2**

*"Clostridium difficile vaccine"*

Introduction

- 5       The invention relates to vaccines to provide immunological protection against *C. difficile* infection.

Background

- 10       *Clostridium difficile* is a common nosocomial pathogen and a major cause of morbidity and mortality among hospitalised patients throughout the world [Kelly et al., 1994]. Outbreaks of *C. difficile* have necessitated ward and partial hospital closure. With the increasing elderly population and the changing demographics of the population, *C. difficile* is set to become a major problem in the 21st century. The
- 15       spectrum of *C. difficile* diseases range from asymptomatic carriage to mild diarrhoea to fulminant pseudomembranous colitis. Host factors rather than bacterial factors appear to determine the response to *C. difficile* [Cheng et al., 1997; McFarland et al., 1991; Shim et al., 1998].
- 20       Reports indicate that hypogammaglobulinaemia in children appears to predispose to the development of disease due to *C. difficile* and that therapy with intravenously administered gamma globulin can be associated with the clinical resolution of chronic relapsing colitis due to *C. difficile* disease [Leung et al., 1991; Pelmutter et al., 1985]. A study by Mulligan et al. [1993] found elevated levels of
- 25       immunoglobulins reactive with *C. difficile* in asymptomatic carriers as opposed to symptomatic patients. Recently it has been shown that patients who became colonised with *C. difficile* who had relatively low levels of serum IgG antibody against toxin A had a much greater risk of developing *C. difficile* diarrhoea [Kyne et al., 2000].
- 30       It is clear that any advance in the understanding of *C. difficile* disease and methods of preventing or treating *C. difficile* diarrhoea (CDD) and other related diseases will be of major therapeutic potential.

Statements of Invention

According to the invention there is provided a vaccine for the treatment or prophylaxis of *C. difficile* associated disease, the vaccine comprising a *C. difficile* gene or a *C. difficile* peptide/polypeptide or a derivative or fragment or mutant or variant thereof which is immunogenic in humans.

The invention also provides a vaccine for the treatment or prophylaxis of *C. difficile* associated disease, the vaccine comprising a *C. difficile* gene or *C. difficile* peptide/polypeptide or a derivative or fragment or mutant or variant thereof to which immunoreactivity is detected in individuals who have recovered from *C. difficile* infection.

Preferably the gene encodes a *C. difficile* surface layer protein, SlpA or variant or homologue thereof.

Preferably the peptide/polypeptide is a *C. difficile* surface layer protein, SlpA or variant or homologue thereof.

Most preferably the vaccine comprises a chimeric nucleic acid sequence. Preferably the chimeric nucleic acid sequence is derived from the 5' end of the gene, encoding the mature N-terminal moiety of SlpA from *C. difficile*.

In one embodiment of the invention the vaccine comprises a chimeric peptide/polypeptide. Preferably the amino acid sequence of the chimeric peptide/polypeptide is derived from the mature N-terminal moiety of SlpA from *C. difficile*.

Preferably the vaccine of the invention contains an amino acid sequence SEQ ID No.1 or a derivative or fragment or mutant or variant thereof.

Preferably the vaccine contains an amino acid sequence SEQ ID No.2 or a derivative or fragment or mutant or variant thereof.

5 In one embodiment of the invention the vaccine contains a nucleotide sequence SEQ ID No.3 or a derivative or fragment or mutant or variant thereof; a nucleotide sequence SEQ ID No.4 or a derivative or fragment or mutant or variant thereof; a nucleotide sequence SEQ ID No.5 or a derivative or fragment or mutant or variant thereof; a nucleotide sequence SEQ ID No.6 or a derivative or fragment or mutant or variant thereof; a nucleotide sequence SEQ ID No.7 or a derivative or fragment or mutant or variant thereof; a nucleotide sequence SEQ ID No.8 or a derivative or fragment or mutant or variant thereof; a nucleotide sequence SEQ ID No.9 or a derivative or fragment or mutant or variant thereof or a nucleotide sequence SEQ ID No.10 or a derivative or fragment or mutant or variant thereof.

15 Preferably the vaccine of the invention is in combination with at least one other *C. difficile* sub-unit.

The invention provides a vaccine for the treatment or prophylaxis of *C. difficile* associated disease, the vaccine comprising the mature N-terminal moiety of a surface layer protein, SlpA of *C. difficile* or variant or homologue thereof which is immunogenic in humans.

20 Most preferably the N-terminal moiety of SlpA contains an amino acid sequence SEQ ID No. 1.

25 In one embodiment of the invention the N-terminal moiety of SlpA contains an amino acid sequence SEQ ID No. 2.

30 The invention also provides a vaccine for the treatment or prophylaxis of *C. difficile* associated disease, the vaccine comprising an immunodominant epitope derived

from a *C. difficile* gene or a *C. difficile* peptide/polypeptide or a derivative or fragment or mutant or variant thereof which is immunogenic in humans.

5 Preferably the vaccine of the invention comprises a pharmaceutically acceptable carrier. Most preferably the vaccine is in combination with a pharmacologically suitable adjuvant. Ideally the adjuvant is interleukin 12. Alternatively the adjuvant may be a heat shock protein.

10 In one embodiment of the invention the vaccine comprises at least one other pharmaceutical product.

The pharmaceutical product may be an antibiotic, selected from one or more metronidazole, amoxycillin, tetracycline or erythromycin, clarithromycin or tinidazole.

15 In one embodiment of the invention the pharmaceutical product comprises an acid-suppressing agent such as omeprazole or bismuth salts.

20 The vaccine of the invention may be in a form for oral administration, intranasal administration, intravenous administration or intramuscular administration.

In one embodiment of the invention the vaccine includes a peptide delivery system.

25 The invention also provides an immunodominant epitope derived from a *C. difficile* gene or a *C. difficile* peptide/polypeptide or a derivative or fragment or mutant or variant thereof. Preferably the *C. difficile* peptide/polypeptide contains an amino acid sequence SEQ ID No.1 or SEQ ID No.2 or a derivative or fragment or mutant or variant thereof.

30 In one embodiment of the invention the *C. difficile* peptide/polypeptide contains an amino acid sequence SEQ ID No.3 or SEQ ID No.4 or SEQ ID No.5 or SEQ ID No.6 or SEQ ID No.7 or SEQ ID No.8 or SEQ ID No. 9 or SEQ ID No. 10 or a derivative or fragment or mutant or variant thereof.

The invention further provides a chimeric nucleic acid sequence derived from the 5' end of the *slpA* gene encoding the mature N-terminal moiety of SlpA from *C. difficile* which is immunogenic in humans.

5 The invention also provides a chimeric peptide/polypeptide wherein the amino acid sequence of the chimeric peptide/polypeptide is derived from the mature N-terminal moiety of SlpA from *C. difficile*.

10 The invention provides a *C. difficile* peptide comprising SEQ ID No. 1 or SEQ ID No. 2 or SEQ ID No. 3 or SEQ ID No. 4 or SEQ ID No. 5 or SEQ ID No. 6 or SEQ ID No. 7 or SEQ ID No. 8 or SEQ ID No. 9 or SEQ ID No. 10.

15 One aspect of the invention provides for the use of a *C. difficile* gene or a *C. difficile* peptide/polypeptide or a derivative or fragment or mutant or variant thereof which is immunogenic in humans in the preparation of a medicament for use in a method for the treatment or prophylaxis of *C. difficile* infection or *C. difficile* associated disease in a host.

20 Preferably the medicament which is prepared is a vaccine of the invention.

The invention also provides a method for preparing a vaccine for prophylaxis or treatment of *C. difficile* associated disease, the method comprising;

25 obtaining a *C. difficile* gene or a *C. difficile* peptide/polypeptide or a derivative or fragment or mutant or variant thereof which is immunogenic in humans; and

30 forming a vaccine preparation comprised of said gene or peptide/polypeptide or derivative or fragment or mutant or variant, which is suitable for administration to a host and which when administered raises an immune response.

Preferably the *C. difficile* peptide/polypeptide contains an amino acid sequence SEQ ID No.1 or SEQ ID No.2 or a derivative or fragment or mutant or variant thereof.

5 Most preferably the *C. difficile* gene contains an amino acid sequence SEQ ID No.3 or SEQ ID No.4 or SEQ ID No.5 or SEQ ID No.6 or SEQ ID No.7 or SEQ ID No.8 or SEQ ID No.9 or SEQ ID No.10 or a derivative or fragment or mutant or variant thereof.

10 The invention further provides a method for prophylaxis or treatment of *C. difficile* associated disease, the method comprising;

obtaining a *C. difficile* gene or a *C. difficile* peptide/polypeptide or a derivative or fragment or mutant or variant thereof which is immunogenic in humans;

15 forming a vaccine preparation comprised of said gene or peptide/polypeptide or derivative or fragment or mutant or variant, and

administering the vaccine preparation to a host to raise an immune response.

20 One aspect of the invention provides monoclonal or polyclonal antibodies or fragments thereof, to a *C. difficile* peptide/polypeptide or a derivative or fragment or mutant or variant thereof which is immunogenic in humans.

25 Another aspect of the invention provides monoclonal or polyclonal antibodies or fragments thereof, to *C. difficile* peptide/polypeptide or a derivative or fragment or mutant or variant thereof to which immunoreactivity is detected in individuals who have recovered from *C. difficile* infection.

30 The invention also provides purified antibodies or serum obtained by immunisation of an animal with a vaccine of the invention.

The invention provides the use of the antibodies or fragments of the invention in the preparation of a medicament for treatment or prophylaxis of *C. difficile* infection or *C. difficile* associated disease.

5 Preferably the antibodies or serum are used in the preparation of a medicament for treatment or prophylaxis of *C. difficile* infection or *C. difficile* associated disease.

Most preferably the antibodies or fragments or serum of the invention are used in passive immunotherapy for established *C. difficile* infection.

10

In one embodiment of the invention the antibodies or fragment or serum of the invention are used for the eradication of *C. difficile* associated disease.

15

The invention also provides use of interleukin 12 as an adjuvant in *C. difficile* vaccine.

The invention further provides use of humanised antibodies or serum for passive vaccination of an individual with *C. difficile* infection.

20

#### Brief Description of the Drawings

25

The invention will be more clearly understood from the following description thereof given by way of example only with reference to the accompanying figures, in which:-

30

Fig. 1A is a Western blot showing recognition of antigens from a crude extract of *C. difficile* 171500 (PCR type 1) by serum antibodies from a patient infected with this strain. Lane 1: Pre-infection; Lane 2: Early acute; Lane 3: Late acute; Lane 4: Convalescent;



Fig. 1B is a Western blot showing recognition of antigens from a crude extract of *C. difficile* 170324 (PCR type 12) by serum antibodies from a patient infected with this strain. Lane 1: Pre-infection; Lanes 2-5: Acute; Lanes 6-7: Convalescent;

5

Fig. 2. is a Western blot showing recognition of antigens from two *C. difficile* strains of different type by serum from convalescent patients.

Lane 1: Strain 170324 (PCR type 12), crude antigen preparation

Lane 2: Strain 170324, surface layer protein preparation

10

Lane 3: Strain 171500 (PCR type 1), crude antigen preparation

Lane 4: Strain 171500, surface layer protein preparation.

Molecular mass markers (kDa) are shown on the left; and

15

Fig. 3 is an SDS-PAGE gel showing crude SLP preparations from selected strains of *C. difficile*. The gel contains 12% acrylamide, and has been stained for protein with Coomassie Blue. Each lane contains 5 µg of protein. Molecular weight markers are shown on the left.

20

Lane 1: 171500 (PCR type 1)

Lane 2: 172450 (PCR type 5)

Lane 3: 170324 (PCR type 12)

Lane 4: 171448 (PCR type 12)

Lane 5: 171862 (PCR type 17)

Lane 6: 173644 (PCR type 31)

25

Lane 7: 170444 (PCR type 46)

Lane 8: 170426 (PCR type 92)

#### Detailed Description of the invention

30

Two antigenic peptides containing SEQ ID No. 1 and SEQ ID No. 2, associated with two common infecting types of *C. difficile*, were found to be immunogenic in humans. The antigenic peptides were found to induce a strong immune response in

individuals who recover from *C. difficile* infection. Individuals who have recovered from *C. difficile* infection are those individuals who have been exposed to *C. difficile* or something strongly related and have recovered. This includes individuals where a carrier state exists in that the *C. difficile* infection has not and will not necessarily become clinically significant.

These antigenic peptides were found to be products of the *slpA* gene from *C. difficile* which is the structural gene for the surface layer protein, SlpA. The gene or its products are therefore ideal candidates for the preparation of vaccines against *C. difficile*.

Surface layer proteins (SLPs), also known as S-layers or crystalline surface layers, are associated with a wide range of bacterial species. They form a 2-dimensional array, which covers the surface of the cell completely, and grows with the cell [Sleytr et al., 1993]. The molecular weight can range from 40 000 to 200 000 Da. The proteins are typically acidic, contain a large proportion of hydrophobic amino acid residues, and have few or no sulphur-containing amino acid residues. Glycosylated S-layer proteins occur in some species. The precise function of S-layers is not always known, but since they comprise approximately 15% of the cell protein, it seems likely that they are important for *in vivo* functioning of the organism. In Gram positive organisms, the SLP has been shown to delay or prevent the excretion of degradative enzymes from the cell to the outside milieu, and may thereby create a space analogous to the periplasmic space of Gram negative bacteria. Many pathogenic species possess SLPs, which have been ascribed functions such as antiphagocytosis (*Campylobacter fetus*), and inhibition of complement-mediated killing (*Aeromonas salmonicida*).

Kawata et al. [1984] described the SLPs of *Clostridium difficile*. They showed the S-layer to be composed of 2 polypeptides, and demonstrated size heterogeneity for the polypeptides from different strains. Delmée et al. [1986] showed that crude extracts from *C. difficile* strains of different serotype showed different polypeptide profiles in SDS-PAGE. Poxton et al. [1999] made similar observations using purified SLP preparations. Slide agglutination [Delmée et al., 1990] has identified 21 different serotypes, apparently distinguished by the heterogeneity of the SLP.

35

Pantosti et al. [1989] isolated *C. difficile* from a number of patients with antibiotic-associated diarrhoea, and prepared SLPs from them.. Cerquetti et al. [2000] published N-terminal sequences of SLPs from several strains, indicating wide differences between strains.. In 2000 the complete DNA sequence of the *C. difficile* genome was published (available at web address [http://www.sanger.ac.uk/Projects/C\\_difficile/](http://www.sanger.ac.uk/Projects/C_difficile/)).

The peptides of the invention were found to be encoded by a single open reading frame (ORF) named *slpA* from *C. difficile*. The peptides identified in our clinical study correspond to a lower molecular weight moiety of the *slpA* gene product. Since an immune response is also mounted against a higher molecular weight *slpA* gene product (Fig. 2), this entity may also be included in a vaccine.

The *slpA* gene has been sequenced from a number of strains corresponding to different PCR types. The sequences of strains 171500 (PCR type 1)(NCIMB 41081; PHLS R13537), 172450 (PCR type 5)(PHLS R12884), 170324 (PCR type 12) (NCIMB 41080; PHLS R12882), 171448 (PCR type 12) (PHLS R13550), 171862 (PCR type 17) (PHLS R13702), 173644 (PCR type 31) (PHLS R13711), 170444 (PCR type 46) (PHLS R12883) and 170426 (PCR type 92) (PHLS R12871) with translations thereof are given in Appendices 1 to 8. Substantial variation in nucleotide and predicted amino acid sequence was found between strains of PCR types 1, 5, 12, 17 and 31. The genes from strains of PCR types 46 and 92 are almost identical in sequence to those of PCR type 12. When the DNA sequences of genes of different strains within a PCR type are compared, the sequences are almost if not quite identical, indicating that the potential for variation is not infinite. These findings are in agreement with serotyping studies [Delmée et al., 1986, 1990], and indicate that the production of an effective vaccine based on the *slpA* product is feasible. In this respect, the present invention includes all variant *slpA* genes and their products, individually and combined, fragments of them, and their mutants and derivatives.

One aspect of the invention provides the combination of immunodominant eptopes from the *slpA* gene products from various serotypes into a single vaccine. In this way a single vaccine may be used to immunise against several different *C. difficile* strains.

The most common PCR types isolated from infections in the clinical study carried out at St. James's Hospital, Dublin, Ireland were PCR types 1 and 12. However, a vaccine which elicits an intense antibody response against many infecting types would be therapeutically very valuable. Recombinant DNA chimera, or several  
5 chimeras, encoding contiguous immunodominant epitopes may be made for use in the vaccine. The recombinant DNA may serve as the active component in a vaccine, or may be inserted into an appropriate expression system for the generation of a chimeric peptide vaccine in a suitable host.

10 Chimeras can be generated by PCR amplification of the DNA encoding peptide regions of interest, incorporating cleavage sites for restriction endonucleases into the primers. The amplified fragments can thus be cleaved to generate compatible ends, and spliced together to create chimeras.

15 The dominant epitopes may be identified by cleavage of the *slpA* products into fragments by agents which cleave at known sites, and by immunoblotting with homologous patient serum. Immunodominant peptides may be tested for their capacity to stimulate T-cell proliferative responses *in vitro*, using mouse splenic T-cells.

20 DNA vaccination involves immunisation with recombinant DNA encoding the antigen or epitope of interest, cloned in a vector which promotes high level expression in mammalian cells. Typically, the vector is a plasmid vector which which also replicates in a procaryotic vector such as *Escherichia coli*, so that the  
25 DNA can be produced in quantity. Following immunisation, the plasmid enters a host cell, where it remains in the nucleus, and directs synthesis of the recombinant polypeptide. The polypeptide stimulates the production of neutralising antibodies, as well as activating cytotoxic T-cells.

30 Using a DNA vaccine, it may be necessary to modify the DNA sequence to take account of codon usage in humans. The G+C content of mammalian DNA is much higher than that of *C. difficile*. The generation of such synthetic DNA molecules, essentially containing numerous silent mutations, is within the scope of the invention.

A peptide vaccine will ideally be made using recombinant peptides. Similar considerations apply as in the generation of a DNA vaccine with regard to expression in a different host, such as *Escherichia coli*, which has a different codon usage pattern to *C. difficile*. Problems of expression may be overcome by the use of a special host strain which carries additional copies of rare tRNAs (e.g. *E. coli* BL21-CodonPlus™-RIL from Stratagene), or by using *de novo* synthesis of a DNA segment carrying silent mutations which will enable normal expression in *E. coli*. There are many expression systems which are likely to allow high-level expression of *slpA* genes in *E. coli*. An example is the pBAD/Thio TOPO vector of Invitrogen, in which expressed genes are under control of the arabinose promoter, which is subject to positive and negative control, enabling very tight control of expression. In this vector, the recombinant protein is typically fused to a modified thioredoxin carrying several histidine residues which enable purification by nickel chromatography. The recombinant protein can be cleaved from the thioredoxin moiety by enterokinase enzyme.

Affinity chromatography may also be used with fixed antibodies or some other agent which strongly binds the peptide of interest to purify the protein from the native organism.

Purified immunogenic peptides may be used in combination with other *C. difficile* sub-units as a combined vaccine against *C. difficile*. Potential candidates are the products of the other *slp* genes, which share limited homology with the *slpA* gene product and with the N-acetylmuramoyl L-alanine amidase, (CwlB), from *Bacillus subtilis*, and which may be involved in remodelling of the peptidoglycan.

Other purified proteins of *C. difficile* to which constitutive antibodies are detected in individuals recovering from *C. difficile* infection are also within the scope of the present invention

A deposit of *Clostridium difficile* strain 171500, PCR type 1, was made at the NCIMB on January 29, 2001, and accorded the accession number NCIMB 41081.

A deposit of *Clostridium difficile* strain 170324, PCR type 12, was made at the NCIMB on January 29, 2001, and accorded the accession number NCIMB 41080.

Two peptides of the invention were found to contain the following sequences:

33kDa peptide

SEQ ID No. 1: DKTKVETADQGYTVVQSKYK

31kDa peptide

SEQ ID No. 2 ATTGTQGYTVVKNDGKKAVK

The invention will be more clearly understood from the following examples.

Example 1. Clinical Study

Examination of sequential antibody responses to *C. difficile* among elderly patients who developed the disease was carried out. The study was based on the hypothesis that the host immune response influenced the development of *Clostridium difficile* disease. In particular we determined that a particular pattern of immune response to *C. difficile* antigens correlated with the outcome of CDD.

Materials and Methods

Patients

Serum was collected from over 300 patients and of these 30 patients developed CDD. The infecting strain (homologous strain) was grown from each patient. Strains of *C. difficile* were typed at the Anaerobe Reference Laboratory, Wales [O'Neill et al., 1996]. The most common strains isolated were PCR type 1 (n = 15) which is the most common type causing epidemics and PCR type 12 (n = 5) which is also a common hospital strain. Pre-infection serum samples were obtained from patients. Acute phase sera were then collected from patients who developed *C. difficile* disease. Convalescent sera were collected from patients who recovered. Protein extracts of patients' infecting *C. difficile* strain were probed with the patients sera using Western blotting. IgG responses to the antigens were examined.

Western blotting

Proteins from SDS-PAGE gels were electroblotted (0.8mA/cm<sup>2</sup> for 1 h) to PVDF membrane using a semi-dry blotting apparatus (Atto). Primary antibodies (human

serum: 1/50 – 1/10,000 dilution) were detected using a 1/5000 dilution of anti-human IgG (horse radish peroxidase-conjugated) in combination with enhanced chemiluminescence (ECL). Blots were washed in phosphate buffered saline (pH 7.5) containing Tween 20 (0.1% v/v), and incubated in the same solution comprising dried skim milk (5% w/v) and antibodies at the appropriate concentration. Blots were exposed to Kodak X-OMAT film for various periods of time and developed.

### Results

Overall 5 patients made a full recovery and new antibody responses to previously unrecognised antigens were evident in 4 of these patients. Three of these patients had *C. difficile* belonging to PCR type 1 and one patient had *C. difficile* PCR type 12. These patients developed an acute phase antibody response to previously unrecognised *C. difficile* antigens which persisted during convalescence (Figs. 1A and 1B). These antigens were recognised by antibodies from patients who recovered and represent potential candidate vaccine antigens. Fig 1A shows a strong reaction of convalescent antibodies was observed with the 33 kDa antigen (Lane 4, arrow). Fig 1B shows a strong reaction of convalescent antibodies was observed with the 31 kDa antigen (Lanes 6 and 7, arrow).

These antibody responses have also been found in some controls in the same ward who were also on antibiotics but who did not develop CDD.

### Example 2. Further characterisation of protective antigens

#### Materials and Methods

Partial purification and N-terminal sequencing of the 33 kDa and the 31 kDa proteins

The antigens were partially purified from *C. difficile* based on their molecular weight using preparative continuous-elution SDS-PAGE on a model 491 Prep-Cell (Bio-Rad). The appropriate antigens were subsequently identified on Western blots probed with serum obtained from individuals who recovered from *C. difficile* infection.

Preparation of surface layer proteins (SLPs)

SLPs were purified from *C. difficile* by extracting washed cells with 8 M urea, in 50 mM Tris HCl, pH 8.3 in the presence of a cocktail of protease inhibitors

(Complete®, Boehringer Mannheim), for 1 h at 37°C, followed by centrifugation for 19 000 x g for 30 min. The SLPs were recovered in the supernatant and dialysed to remove the urea [Cerquetti et al., 2000].

## 5      Results

The immunodominant protein which was associated with a positive outcome from *C. difficile* strain 171500 (PCR type 1) was identified and purified using preparative SDS-PAGE. The N-terminal region of the protein was sequenced using an Applied Biosystems Procise Sequencer, viz DTKKVETADQGYTVVQSKYK (SEQ ID  
10      No. 1)

The antigen which was associated with a protective antibody response from the *C. difficile* strain 170324 (PCR type 12) was identified and the N-terminal sequence obtained, viz ATTGTQGYTVVKNDGKKAVK (SEQ ID No. 2).

15      These sequences were used to interrogate the *C. difficile* genome sequence using the TBLASTN programme, which compared our query sequences with those of the genome project (available at web address  
20      [http://www.sanger.ac.uk/Projects/C\\_difficile/](http://www.sanger.ac.uk/Projects/C_difficile/)), translated in all 6 possible reading frames. A nearly identical stretch of sequence was identified when the sequence from strain 1710324 (type 12) was used for interrogation. The same stretch of sequence was picked up with the sequence from strain 171500 (type 1) was used, although the identity was much less strong. Since the homologous sequence belonged to an open reading frame encoding a 719-residue peptide, this result was  
25      somewhat surprising. However, when the N-terminal sequences from the higher molecular weight SLP component were later published by Cerquetti et al [2000], it became apparent that they were encoded downstream along the same gene, subsequently identified as *slpA*, and the reason for the discrepancy in size between the gene and its products became readily apparent.

30      The purified SLPs from strains 171500 (PCR type 1) and 170324 (PCR type 12) showed strong reactivity with homologous convalescent serum, and co-migrated with the dominant antigens detected in crude cell extracts as shown in Fig. 2. Lanes 1 and 3 contain crude antigen preparations from PCR types 1 and 12 respectively,  
35      and Lanes 2 and 4 contain SLP preparations from PCR types 1 and 12, respectively.



Panel A was probed with serum from a patient recovering from infection with PCR type 1, and Panel B was probed with serum from a patient recovering from infection with PCR type 12. Each serum detected 2 major antigens in the infecting strain (Panel A, Lane 3); (Panel B, Lane 1), which co-migrated with the 2 SLPs (Panel A, Lane 4; Panel B, Lane 2), with which the sera also reacted strongly. Note that serum from the patient infected with the PCR type 1 strain recognised the higher molecular weight SLP from the PCR type 12 strain (Panel A, Lanes 1 and 2), whereas the converse did not occur (Panel B, Lanes 3 and 4). There is no apparent antigenic cross-reactivity with regard to the lower molecular weight SLPs.

SLPs were prepared from selected strains by urea extraction, and subjected to SDS-PAGE and staining with Coomassie Blue (Fig. 3). Most strains showed a characteristic profile, with two major bands located in the 29 000 to 36 000 and 45 000 to 50 000 molecular weight range. An exception was strain 172450 (Fig. 3, Lane 2), which showed a single, high molecular weight band, approximately 43 000 in size.

#### Cloning, sequencing and analysis of *slpA* genes

The nucleotide sequences of the *slpA* genes from the two sample strains of *C. difficile* (PCR types 1 and 12, deposited at the NCIMB) and of several others (PCR types 5, 12, 17, 31, 46 and 92, available from the Anaerobe Reference Unit at the Department of Medical Microbiology and Public Health Laboratory, Cardiff, Wales) were obtained. The *slpA* gene and flanking sequence was amplified by polymerase chain reaction from genomic DNA prepared from *C. difficile* using a commercial kit (Puregene® DNA isolation kit for yeast and Gram positive bacteria, Gentra systems Minneapolis, MN). The forward primer (5' ATGGATTATTATAGAGATGTGAG 3'), was based on sequence from the genome sequencing project, starting 112 nucleotides upstream from the start of the *slpA* open reading frame. Two reverse primers were used, depending on the PCR type. A downstream primer (5' CTATTTAAAGTTTATTAAACTTATATTAC 3') was used to amplify *slpA* from PCR types 12, 17, 31, 46 and 92. A reverse primer based on the 3' end of the *slpA* open reading frame from strain 630 and the subsequent nonsense codon (5' TTACATATCTAATAAATCTTTCATTTTGTTTATAACTG 3') was used to

amplify *slpA* from PCR types 1 and 5. The choice of primer for the latter two PCR types may have resulted in a small number of systematic errors in the nucleotide sequence obtained. PCR was carried out using HotStar™ Taq polymerase (Qiagen Ltd., Crawley, West Sussex, UK) according to the manufacturer's instructions. A single fragment of approximately 2 kb was obtained for each strain, which was then cloned into the pBAD/Thio TOPO vector (Invitrogen, Groningen, Netherlands). Inserts were sequenced from both ends by standard procedures in commercial facilities at MWG (Wolverton Mill South, Milton Keynes, UK) and Cambridge University. New primers were designed on the basis of initial sequencing results, enabling sequencing of both strands to be completed (a process known as chromosome walking).

The results are shown in Appendices 1-8.

The nucleotide sequences were translated to enable prediction of the amino acid sequence(s) of the product(s) (Appendices 1-8). The N-terminal sequences obtained experimentally for the low molecular weight protective antigens from strains 171500 (PCR type 1) and 170324 (PCR type 12) were almost identical to those predicted from the nucleotide sequences of their respective *slpA* genes (18/20 identical residues for strain 171500, and 19/20 identical residues for strain 170324).

Appendix 1 shows the open reading frame with translation for *slpA* from strain 171500 (PCR type 1), SEQ ID No 3. Since the reverse primer was based on the 35 nucleotides from the 3' end of the *slpA* gene, the sequence is not necessarily 100% accurate in this region. However, this part of the gene does not seem to vary greatly from strain to strain.

Appendix 2 shows the open reading frame with translation for *slpA* from strain 172450 (PCR type 5), SEQ ID No 4. Again, the sequence obtained for the 3' 35 nucleotides is not fully reliable. This gene is considerably smaller than the other *slpA* genes sequenced, and shows strong sequence divergence from the other PCR types examined.

Appendix 3 shows the open reading frame with translation for *slpA* from strain 170324 (PCR type 12), SEQ ID No 5. This gene showed a single base difference

when compared with the strain used for the genome sequencing project, strain 630, of the same PCR type. The deduced amino acid sequence is identical.

5 Appendix 4 shows the open reading frame with translation for *slpA* from strain 171448 (PCR type 12), SEQ ID No 6. This gene was almost identical in sequence to that from strain 170324.

10 Appendix 5 shows the open reading frame with translation for *slpA* from strain 171862 (PCR type 17), SEQ ID No 7.

Appendix 6 shows the open reading frame with translation for *slpA* from strain 173644 (PCR type 31), SEQ ID No 8. Like the *slpA* from strain 172450, this sequence is very dissimilar to those of *slpA* genes from other PCR types encountered.

15 Appendix 7 shows the open reading frame with translation for *slpA* from strain 170444 (PCR type 46), SEQ ID No 9. This sequence is virtually identical to that obtained for *slpA* from PCR type 12 and 92 strains.

20 Appendix 8 shows the open reading frame with translation for *slpA* from strain 170426 (PCR type 92), SEQ ID No 10. This sequence is virtually identical to that obtained for *slpA* from PCR type 12 and 46.

25 The cleavage site of the putative signal sequences from both genes was determined from experimental evidence (the N-terminal sequence of the mature proteins as determined by Edman degradation), and by the prediction tool of the Centre for Biological Sequence Analysis at the Technical University of Denmark [Nielsen et al., 1997]. The site for cleavage of the *slpA* gene product to form the mature SLPs was predicted from experimental [Cerquetti et al., 2000, Karjalainen et al., 2001 and  
30 Calabi et al., 2001]. The cleavage site is typically preceded by the motif TKS. However, the relevant motif is likely to be TKG in strain 173644 (PCR type 31). No obvious motif appeared for strain 172450 (PCR type 5). However, the protein produced by type 5 strains does appear to be cleaved; hence we predicted the site to

occur at a point where the SLP sequence aligns with the cleavage sites of other PCR types.

The molecular weight and isoelectric point was calculated for each of the predicted mature proteins by the ExPASy server of the Swiss Institute for Bioinformatics (Table 1). In general, the calculated molecular weights were in fair agreement with apparent molecular masses determined from migration in gels (Fig. 3). No lower molecular weight band was apparent for Strain 172450 (PCR type 5; Lane 2). However, a higher molecular weight band is present, which is similar in size to the predicted weight for the C-terminal moiety. We observed a similar profile for another type 5 strain. It is possible that the lower molecular weight species is subject to degradation in this strain. Another possibility is that it is heavily glycosylated, which can affect staining. All peptides had a predicted isoelectric point below 7, typical of acidic proteins, and characteristic of SLPs in general [Sleyter et al, 1993].

Table 1

<i>C. difficile</i> strain (PCR type)	pI (N-terminal)	pI (C-terminal)	MW (N-terminal)	MW (C-terminal)
171500 (Type 1)	4.83	4.66	33365.41	44220.37
172450 (Type 5)	4.86	4.65	19364.46	42757.63
170324 (Type 12)	4.92	4.58	34228.25	39522.24
171448 (Type 12)	4.98	4.58	34156.18	39492.21
171862 (Type 17)	5.09	4.53	33783.73	39407.11
173644 (Type 31)	5.05	4.56	33626.48	41821.69
170444 (Type 46)	5.06	4.58	34230.31	39522.24
170426 (Type 92)	4.99	4.58	34242.32	39522.24

The translated nucleotide sequences were compared with published SlpA sequences (EMBL Accession numbers AJ300676, and AJ300677 for examples from PCR types 1, and 17 respectively; strain 630 available from the Sanger Institute for PCR type 12; EMBL Accession number AY004256 for a variant from an unnamed PCR type). The Clustal W alignment programme, which is freely available, was used. Where SlpA sequences from our isolates were compared with those of other strains of the same PCR types, they were found to be nearly or quite identical. This observation

indicates, together with existing knowledge from serotyping, that the number of variants of *slpA* is not infinite, and that natural evolution of the gene is not rapid. Table 2 shows a compilation of homologies, based on amino acid residue identity, for the different translated sequences measured against published sequences. Homologies are compiled for the predicted mature peptides, either combined (Table 2A) or as N-terminal (low molecular weight, less conserved moiety) (Table 2B) and C-terminal (high molecular weight, more conserved) (Table 2C) mature peptides according to predicted cleavage sites. It is clear that the SlpA sequences from strains 172450 (PCR type 5) and 173644 (PCR type 31) are quite distinct particularly with respect to N-terminal region.

Table 2A

Strain.type	630 (type 12)	AJ300676 (type 1)	AJ300677 (type 17)	AY004256 (type unknown)
171500.type1	55.2	99.7	55.4	56.42
172450.type5	49.8	54.0	49.9	47.77
170324.type12	100.0	57.8	81.7	59.77
171448.type12	99.7			
171862.type17	82.3	58.7	100	57.54
173644.type31	57.9	59.2	60.1	56.88
170444.type46	99.6			
170426.type92	99.9			

Table 2B

Strain.type	630 (type 12)	AJ300676 (type 1)	AJ300677 (type 17)	AY004256 (type unknown)
171500.type1	35.4	100	34.5	33.54
172450.type5	31.6	32.2	31.0	24.58
170324.type12	100	34.9	64.6	36.14
171448.type12	99.7			
171862.type17	64.3	34.4	100	31.55
173644.type31	37.5	34.1	41.3	31.86
170444.type46	99.1			
170426.type92	99.7			

Table 2C

Strain.type	630 (type 12)	AJ300676 (type 1)	AJ300677 (type 17)	AY004256 (type unknown)
171500.type1	70.2	99.5	71.2	73.80
172450.type5	58.4	60.4	63.0	57.60
170324.type12	100	77.3	97.1	80.00
171448.type12	99.7			
171862.type17	97.3	78.8	100	79.62
173644.type31	74.1	78.9	75.1	75.38
170444.type46	100			
170426.type92	100			

5

The term antibody used throughout the specification includes but is not limited to polyclonal, monoclonal, chimeric, single chain, Fab fragments and fragments produced by a Fab expression library.

10

The antibodies and fragments thereof may be humanised antibodies. Neutralising antibodies such as those which inhibit biological activity of the substance amino acid sequence are especially preferred for diagnostics and therapeutics.

15

Antibodies both polyclonal and monoclonal which are directed against epitopes obtainable from a polypeptide or peptide of the present invention are particularly useful in diagnosis and those which are neutralising are useful in passive immunotherapy.

20

Antibodies may be produced by any of the standard techniques well known in the art.

25

A therapeutically effective amount of the polypeptide, polynucleotide, peptide or antibody of the invention in the form of pharmaceutical composition may be administered. The composition may optionally comprise a pharmaceutically acceptable carrier, diluent or excipients and including combinations thereof. The pharmaceutical composition may be used in conjugation with one or more additional pharmaceutically active compounds and/or adjuvants.

Different adjuvants depending on the host may be used to increase immunological response. The adjuvant may be selected from the group comprising Freund's, mineral gels such as aluminium hydroxide and surface active substances.

5 The vaccine of the invention may be in the form of an immune modulating composition or pharmaceutical composition and may be administered by a number of different routes such as by injection (which includes parenteral, subcutaneous and intramuscular injection) intranasal, intramuscular, mucosal, oral, intra-vaginal, urethral or ocular administration. There may be different formulation/composition  
10 requirements dependent on the different delivery systems.

15 The invention is not limited to the embodiments hereinbefore described which may be varied in detail.

## References

- Calabi E., Ward S., Wren B., Paxton T., Panico M., Morris H., Dell A., Dougan G., Fairweather N. (2001). Molecular characterization of the surface layer proteins from *Clostridium difficile*. Mol. Microbiol. 40:1187-1199.
- 5 Cerquetti M., Molinari A., Sebastianelli A., Diociaiuti M., Petruzzelli R., Cap C., Mastrantonio P. (2000). Characterization of surface layer proteins from different *Clostridium difficile* clinical isolates. Microbial Pathogenesis, 28:363-372.
- Cheng S.H, Lu J.J, Young T.G, Perng C.L, Chi W.M. (1997) *Clostridium difficile*-associated diseases: comparison of symptomatic infection versus carriage on the basis of risk factors, toxin production, and genotyping results. Clin Infect Dis ; 25: 157-8.
- 10 Delmée M., Laroche Y., Avesani V., Cornelis G. (1986). Comparison of serogrouping and polyacrylamide gel electrophoresis for typing *Clostridium difficile*. J. Clin. Microbiol. 24:991-994.
- 15 Delmée M., Avesani V., Delferrière N., Burtonboy G. (1990). Characterization of flagella of *Clostridium difficile* and their role in serogrouping reactions.
- Karjalainen T., Waligora-Dupriet A.-J., Cerquetti M., Spigaglia P., Maggioni A., Mauri P., Mastrantonio P., (2001). Molecular and genomic analysis of genes encoding surface-anchored proteins from *Clostridium difficile*. Infect. Immun. 69:3442-3446.
- 20 Kawata T., Takeoka A., Takumi K., Masuda K. (1984). Demonstration and preliminary characterization of a regular array in the cell wall of *Clostridium difficile*. FEMS Microbiol. Lett 24:323-328.
- Kelly, C.P., Pothoulakis C and LaMont J.T. *Clostridium difficile* colitis. New England Journal of Medicine. 1994 330: 257-262.
- 25 Kyne L, Warny M, Qamar A, Kelly C. Asymptomatic carriage of *Clostridium difficile* and serum levels of IgG antibody against Toxin A. New England Journal of Medicine 2000; 390-7.
- Leung Y.M, Kelly C.P, Boguniewicz M, Pothoulakis C, LaMont J.T, Flores A. Treatment with intravenous gamma globulin of chronic relapsing colitis by *Clostridium difficile*; toxin: J. Pediatr 1991; 118: 633-7.
- 30 McFarland L.V, Elmer G.W, Stamm W.E, Mulligan M.E. Correlation of immunoblot type, enterotoxin production, and cytokine production with clinical manifestation of *Clostridium difficile* infection in a cohort of hospitalised patients. Infect Immun. 1991; 59: 2456-62.



- Mulligan M.E, Miller S.D, McFarland L.V, Fung H.C, Kwok R.Y. Elevated levels of serum immunoglobulins in asymptomatic carriers of *Clostridium difficile*. *Clin Infect Dis* 1993; 16(Suppl 4); S239-44.
- 5      Nielsen H., Engelbrecht J., Brunak S., von Heijne G. (1997). Identification of prokaryotic and eukaryotic signal peptides and prediction of their cleavage sites. *Protein Eng.* 10:1-6.
- O'Neill G.L., Ogunsofola F.T, Brazier J.S, Duerdon B.I, Modification of a PCR ribotyping method for application as a routine typing scheme for *Clostridium difficile*.  
10      *Anaerobe* (1996) 2, 205-209.
- Pantosti A, Cerquetti M, Viti F, Ortisi G, Mastrantonio P. Immunoblot Of Serum Immunoglobulin G Response to Surface Proteins of *Clostridium difficile* in Patients With Antibiotic Associated Diarrhoea. *J. Clin Microbiol* 1989; 27; 2594-7.
- Pelmutter D.H, Leichtner A.M, Goldman H, Winter H.S. Chronic diarrhoea associated with  
15      hypogammaglobulinaemia and enteropathy in infants and children: *Dig Dis Sci* 1985; 30; 1149-55.
- Poxton I.R., Higgins P.G., Currie C.G., McCoubrey J. (1999). Variation in the cell surface proteins of *Clostridium difficile*. *Anaerobe* 5:213-215.
- Shim J. Johnson S, Samone M, Bliss DZ, Gerding D.N. Primary symptomless colonisation  
20      by *Clostridium difficile* and decreased risk of subsequent diarrhoea. *The Lancet* Vol 351 1998: 633-5.
- Sleytr U.B., Messner P., Pum D., Sára M. (1993). Crystalline bacterial cell surface layers. *Mol. Microbiol.* 10:911-916.

## Appendix 1

5 SEQ ID No. 3. Nucleotide sequence of *slpA* from *Clostridium*  
*difficile* strain 171500, PCR type 1, with translation. The putative  
 secretory signal cleavage site ( $\Delta$ ) and site of cleavage to form the  
 two mature SLPs ( $\blacklozenge$ ) are indicated.

```

1  ATGAATAAGAAAAATATAGCAATAGCTATGTCAGGTTTAAACAGTTTTAGCTTCGGCTGCA  60
-----+-----+-----+-----+-----+-----+-----+-----+
10      1  M  N  K  K  N  I  A  I  A  M  S  G  L  T  V  L  A  S  A  A
20
      61
15  CCTGTATTTGCAGATGATACAAAAGTTGAACTGGTGATCAAGGATATACAGTGGTACAA  120
-----+-----+-----+-----+-----+-----+-----+-----+
      +
      21  P  V  F  A  D  D  T  K  V  E  T  G  D  Q  G  Y  T  V  V  Q
40
      121
20  AGCAAGTATAAGAAAGCTGTTGAACAATTACAAAAGGAATATTAGATGGAAGTATAACA  180
-----+-----+-----+-----+-----+-----+-----+-----+
      41  S  K  Y  K  K  A  V  E  Q  L  Q  K  G  I  L  D  G  S  I  T
60
      181
25  GAAATTAAAGTTTTCTTTGAGGGAACCTTTAGCATCTACTATAAAAGTAGGTTCTGAGCTT  240
-----+-----+-----+-----+-----+-----+-----+-----+
      61  E  I  K  V  F  F  E  G  T  L  A  S  T  I  K  V  G  S  E  L
80
      241
30  AATGCAGCAGATGCAAGTAAATTATTGTTTACACAAGTAGATAATAAACTAGATAATTTA  300
-----+-----+-----+-----+-----+-----+-----+-----+
      81  N  A  A  D  A  S  K  L  L  F  T  Q  V  D  N  K  L  D  N  L
100
      301
35  GGTGATGGAGATTATGTAGATTTCTTAATAACTTCTCCAGGTCAAGGGGATAAAATAACT  360
-----+-----+-----+-----+-----+-----+-----+-----+
      101  G  D  G  D  Y  V  D  F  L  I  T  S  P  G  Q  G  D  K  I  T
120
      361
40  ACAAGTAAACTTGTTGCATTGAAAGATTTAACAGGTGCTTCAGCAGATGCTATAATTGCT  420
-----+-----+-----+-----+-----+-----+-----+-----+
      121  T  S  K  L  V  A  L  K  D  L  T  G  A  S  A  D  A  I  I  A
140
      421
45  GGAACATCTTCAGCAGATGGTGTGTTACAAATACTGGAGCTGCTAGTGGTTCTACTGAG  480
-----+-----+-----+-----+-----+-----+-----+-----+
      141  G  T  S  S  A  D  G  V  V  T  N  T  G  A  A  S  G  S  T  E
160

```

481  
ACAAATTCAGCAGGAACAAACTTGCAATGTCAGCTATTTTTGACACAGCATATACAGAT 540  
-----+-----+-----+-----+-----+-----  
161 T N S A G T K L A M S A I F D T A Y T D  
5 180  
541  
TCATCTGAAACTGCGGTAAAGATTACTATAAAAGCAGATATGAATGATACTAAATTTGGT 600  
-----+-----+-----+-----+-----+-----  
181 S S E T A V K I T I K A D M N D T K F G  
10 200  
601  
AAAGCAGGTGAGACAACTTATTCAACTGGGCTTACATTTGAAGATGGGTCTACAGAAAAA 660  
-----+-----+-----+-----+-----+-----  
201 K A G E T T Y S T G L T F E D G S T E K  
15 220  
661  
ATTGTTAAATTAGGGGACAGTGATATTATAGATATAACTAAAGCTCTTAACTTACTGTT 720  
-----+-----+-----+-----+-----+-----  
221 I V K L G D S D I I D I T K A L K L T V  
20 240  
721  
GTTCTGGAAGTAAAGCAACTGTAAAGTTTGTGAAAAACACCAAGTGCCAGTGTTCAA 780  
-----+-----+-----+-----+-----+-----  
241 V P G S K A T V K F A E K T P S A S V Q  
25 260  
781  
CCAGTAATAACAAAGCTTAGAATAATAAATGCTAAAGAAGAAACAATAGATATTGACGCT 840  
-----+-----+-----+-----+-----+-----  
261 P V I T K L R I I N A K E E T I D I D A  
30 280  
841  
AGTTCCTAGTAAAACAGCACAAAGATTTAGCTAAAAAATATGTATTTAATAAACTGATTTA 900  
-----+-----+-----+-----+-----+-----  
281 S S S K T A Q D L A K K Y V F N K T D L  
35 300  
901  
AATACTCTTTATAAAGTATTAAATGGAGATGAAGCAGATACTAATGGATTAATAGAAGAA 960  
-----+-----+-----+-----+-----+-----  
301 N T L Y K V L N G D E A D T N G L I E E  
40 320  
961  
GTTAGTGGAAATATCAAGTAGTTCTTTATCCAGAAGGAAAAAGAGTTACAACCTAAGAGT 1020  
-----+-----+-----+-----+-----+-----  
321 V S G K Y Q V V L Y P E G K R V T T K S  
45 340  
1021  
GCTGCAAAGGCTTCAATTGCTGATGAAAATTCACCAGTTAAATTAACCTCTTAAGTCAGAT 1080  
-----+-----+-----+-----+-----+-----  
341 A A K A S I A D E N S P V K L T L K S D  
50 360  
◆  
1081  
AAGAAGAAAGACTTAAAGATTATGTGGATGATTTAAGAACATATAATAATGGATATTCA 1140  
-----+-----+-----+-----+-----+-----  
361 K K K D L K D Y V D D L R T Y N N G Y S  
55 380

1141  
AATGCTATAGAAGTAGCAGGAGAAGATAGAATAAGTCAAAAA 1200  
-----+-----+-----+-----+-----  
5 381 N A I E V A G E D R I E T A I A L S Q K  
400  
1201  
TATTATAACTCTGATGATGAAAATGCTATATTTAGAGATTGAGTTGATAATGTAGTATTG 1260  
-----+-----+-----+-----+-----  
10 401 Y Y N S D D E N A I F R D S V D N V V L  
420  
1261  
GTTGGAGGAAATGCAATAGTTGATGGACTTGTAGCTTCTCCTTTAGCTTCTGAAAAGAAA 1320  
-----+-----+-----+-----+-----  
15 421 V G G N A I V D G L V A S P L A S E K K  
440  
1321  
GCTCCTTTATTATTAACTTCAAAAGATAAATTAGATTCAAGCGTAAAAGCTGAAATAAAG 1380  
-----+-----+-----+-----+-----  
20 441 A P L L L T S K D K L D S S V K A E I K  
460  
1381  
AGAGTTATGAATATAAAGAGTACAACAGGTATAAATACTTCAAAGAAAGTTTATTAGCT 1440  
-----+-----+-----+-----+-----  
25 461 R V M N I K S T T G I N T S K K V Y L A  
480  
1441  
GGTGGAGTTAATTCTATATCTAAAGAAGTAGAAAATGAATTAAAAGATATGGGACTTAAA 1500  
-----+-----+-----+-----+-----  
30 481 G G V N S I S K E V E N E L K D M G L K  
500  
1501  
GTTACAAGATTAGCAGGAGATGATAGATATGAAACTTCTCTAAAAATAGCTGATGAAGTA 1560  
-----+-----+-----+-----+-----  
35 501 V T R L A G D D R Y E T S L K I A D E V  
520  
1561  
GGTCTTGATAATGATAAAGCATTTGTAGTTGGAGGAACAGGATTAGCAGATGCCATGAGT 1620  
-----+-----+-----+-----+-----  
40 521 G L D N D K A F V V G G T G L A D A M S  
540  
1621  
ATAGCTCCAGTTGCATCTCAATTAAGAAATGCTAATGGTAAAATGGATTTAGCTGATGGT 1680  
-----+-----+-----+-----+-----  
45 541 I A P V A S Q L R N A N G K M D L A D G  
560

1681  
GATGCTACACCAATAGTAGTTGTAGATGGAAAAGCTAAAACTATAAAATGATGATGTAAAA 1740  
-----+-----+-----+-----+-----  
561 D A T P I V V V D G K A K T I N D D V K  
580  
1741  
GATTTCTTAGATGATTCACAAGTTGATATAATAGGTGGAGAAAACAGTGTATCTAAAGAT 1800  
-----+-----+-----+-----+-----  
581 D F L D D S Q V D I I G G E N S V S K D  
600  
1801  
GTTGAAAATGCAATAGATGCTACAGGTAAATCTCCAGATAGATATAGTGGAGATGAT 1860  
-----+-----+-----+-----+-----  
601 V E N A I D D A T G K S P D R Y S G D D  
620  
1861  
AGACAAGCAACTAATGCAAAAGTTATAAAAGAATCTTCTTATTATCAAGATAACTTAAAT 1920  
-----+-----+-----+-----+-----  
621 R Q A T N A K V I K E S S Y Y Q D N L N  
640  
1921  
AATGATAAAAAAGTAGTTAATTTCTTTGTAGCTAAAGATGGTTCTACTAAAGAAGATCAA 1980  
-----+-----+-----+-----+-----  
641 N D K K V V N F F V A K D G S T K E D Q  
660  
1981  
TTAGTTGATGCTTTAGCAGCAGCTCCAGTTGCAGCAAACCTTTGGTGTAACCTCTTAATTCT 2040  
-----+-----+-----+-----+-----  
661 L V D A L A A A P V A A N F G V T L N S  
680  
2041  
GATGGTAAGCCAGTAGATAAAGATGGTAAAGtATTAACCTGGTTCTGATAATGATAAAAAT 2100  
-----+-----+-----+-----+-----  
681 D G K P V D K D G K V L T G S D N D K N  
700  
2101  
AAATTAGTATCTCCAGCACCTATAGTATTAGCTACTGATTCTTTATCTTCAGATCaAAGT 2160  
-----+-----+-----+-----+-----  
701 K L V S P A P I V L A T D S L S S D Q S  
720  
2161  
GTATCTATAAGTAaAGTTCTTGATAAAGATAATGGAGAAAACCTTAGTTCAAGTTGGTAAA 2220  
-----+-----+-----+-----+-----  
721 V S I S K V L D K D N G E N L V Q V G K  
740  
2221 GGTATAGCTACTTCAGTTATAAACAAAATGAAAGATTTATTAGATATG 2268  
-----+-----+-----+-----+-----  
741 G I A T S V I N K M K D L L D M 756

50

## Appendix 2

5 SEQ ID No. 4. Nucleotide sequence of *slpA* from *Clostridium difficile* strain 172450, PCR type 5, with translation. The putative secretory signal cleavage site ( $\Delta$ ) is indicated, and an approximation of the and site of cleavage to form the two mature SLPs ( $\blacklozenge$ ) is also indicated.

```

10      1
      ATGAAAAAAGAAATTTAGCAATGGCTATGGCAGCTGTTACTGTAGTAGGTTCTGCTGCT      60
      -----+-----+-----+-----+-----+-----+-----+-----
      1  M  K  K  R  N  L  A  M  A  M  A  A  V  T  V  V  G  S  A  A
20
      61
15      CCAGTTTTTGCAGCAGCTTCAGATGTAATATCACTACAAGATGGTACAAATGATAAGTAT      120
      -----+-----+-----+-----+-----+-----+-----+-----
      21  P  V  F  A  A  A  S  D  V  I  S  L  Q  D  G  T  N  D  K  Y
40
      121
20      ACAGTATCAAATACTAAAGCTAGTGACTTAGTAAAGGATATTTTAGCAGCACAAAACCTTA      180
      -----+-----+-----+-----+-----+-----+-----+-----
      41  T  V  S  N  T  K  A  S  D  L  V  K  D  I  L  A  A  Q  N  L
60
      181
25      ACAACAGGTGCAGTTATTTTGAACAAAGATACAAAAGTTACTTTCTATGATGCAAATGAG      240
      -----+-----+-----+-----+-----+-----+-----+-----
      61  T  T  G  A  V  I  L  N  K  D  T  K  V  T  F  Y  D  A  N  E
80
      241
30      AAAGATTCTTCAACTCCAAGTGGAGATAAAAAAGTTTATTTCAGAACAACTTTAACTACA      300
      -----+-----+-----+-----+-----+-----+-----+-----
      81  K  D  S  S  T  P  T  G  D  K  K  V  Y  S  E  Q  T  L  T  T
100
      301
35      GCTAATGGAAATGAAGATTATGTAAAGACAACCTTTAAAAAATTTAGATGCAGGAGAATAT      360
      -----+-----+-----+-----+-----+-----+-----+-----
      101 A  N  G  N  E  D  Y  V  K  T  T  L  K  N  L  D  A  G  E  Y
120
      361
40      GCTATTATAGATTTAACTTATAATAATGCTAAAAGTGTGAAATTAAAGTAGTAGCAGCT      420
      -----+-----+-----+-----+-----+-----+-----+-----
      121 A  I  I  D  L  T  Y  N  N  A  K  T  V  E  I  K  V  V  A  A
140
      421
45      AGTGAAAAACAGTAGTTGTATCTAGTGATGCGAAAAATAGTGCAAAAGATATAGCTGAA      480
      -----+-----+-----+-----+-----+-----+-----+-----
      141 S  E  K  T  V  V  V  S  S  D  A  K  N  S  A  K  D  I  A  E
160
      481
50      AAATATGTGTTTGAAGACAAAGACTTAGAAAAATGCACTAAAAACTATAAATGCCTCAGAT      540
      -----+-----+-----+-----+-----+-----+-----+-----
      161 K  Y  V  F  E  D  K  D  L  E  N  A  L  K  T  I  N  A  S  D
55      180

```

541  
TTCAGTAAACTGATAGTTACTATCAAGTAGTTCTTTATCCAAAAGGAAAGAGATTACAA 600  
-----+-----+-----+-----+-----+-----  
181 F S K T D S Y Y Q V V L Y P K G K R L Q  
200  
601  
GGTTTCTCAACTTATAGAGCTACAAATTATAATGAAGGAACTGCATATGGTAATACACCA 660  
-----+-----+-----+-----+-----+-----  
201 G F S T Y R A T N Y N E G T A Y G N T P  
220  
◆  
661  
GTAATATTAAGTCTAAATCTACTAGTAAGAGTAATTTAAAGACTGCAGTAGAAGAGTTA 720  
-----+-----+-----+-----+-----+-----  
221 V I L T L K S T S K S N L K T A V E E L  
240  
721  
CAAAAATTGAATGCTAGTTATTCTAATACTACAACCTTTAGCTGGTGATGACAGAATACAA 780  
-----+-----+-----+-----+-----+-----  
241 Q K L N A S Y S N T T T L A G D D R I Q  
260  
781  
ACAGCTATAGAGATAAGTAAAGAATATTACAATAATGATGGCGAGAAATCAGATCATTCA 840  
-----+-----+-----+-----+-----+-----  
261 T A I E I S K E Y Y N N D G E K S D H S  
280  
841  
GCTGATGTTAAAGAGAATGTTAAAAATGTTGTATTAGTAGGTGCAAATGCACTAGTAGAT 900  
-----+-----+-----+-----+-----+-----  
281 A D V K E N V K N V V L V G A N A L V D  
300  
901  
GGATTAGTTGCGGCTCCTTTAGCAGCAGAAAAAGATGCTCCACTATTATTAAGTTCAAAA 960  
-----+-----+-----+-----+-----+-----  
301 G L V A A P L A A E K D A P L L L T S K  
320  
961  
GATAAATTAGATTTCGTCAGTAAATCTGAAATAAAGAGAGTTTTAGACTTAAAACTTCA 1020  
-----+-----+-----+-----+-----+-----  
321 D K L D S S V K S E I K R V L D L K T S  
340  
1021  
ACAGAAGTAACAGGAAAAACAGTTTATATAGCTGGTGGAGTTAATAGTGTATCTAAAGAA 1080  
-----+-----+-----+-----+-----+-----  
341 T E V T G K T V Y I A G G V N S V S K E  
360  
1081  
GTTGTAACAGAATTAGAATCAATGGGATTAAAAGTTGAAAGATTCTCAGGTGATGATAGA 1140  
-----+-----+-----+-----+-----+-----  
361 V V T E L E S M G L K V E R F S G D D R  
380  
1141  
TATGAACTTCTTTAAAAATAGCAGGTGAAATAGGCTTAGATAATGATAAGGCTTATGTA 1200  
-----+-----+-----+-----+-----+-----

381 Y E T S L K I A G E I G L D N D K A Y V  
400  
1201  
5 GTTGGTGGACAGGATTAGCAGATGCCATGAGTATAGCTTCAGTTGCTTCTACTAAATTA 1260  
-----+-----+-----+-----+-----  
401 V G G T G L A D A M S I A S V A S T K L  
420  
1261  
10 GATGGTAATGGTGTGTAGATAGAACAAATGGACATGCTACTCCAATAGTTGTTGTAGAT 1320  
-----+-----+-----+-----+-----  
421 D G N G V V D R T N G H A T P I V V V D  
440  
1321  
15 GGAAAAGCTGATAAAATATCTGATGACTTAGATAGTTTCTTAGGAAGCGCTGATGTAGAT 1380  
-----+-----+-----+-----+-----  
441 G K A D K I S D D L D S F L G S A D V D  
460  
1381 ATAATAGGTGGATTTGCAAGTGTATCTGAAAAGATGGAAGAAGCTATATCAGATGCTACT  
1440  
20 -----+-----+-----+-----+-----  
461 I I G G F A S V S E K M E E A I S D A T  
480  
1441  
25 GGTAAAGGCGTTACAAGAGTTAAAGGCGACGATAGACAAGACACTAACTCTGAAGTTATA 1500  
-----+-----+-----+-----+-----  
481 G K G V T R V K G D D R Q D T N S E V I  
500  
1501  
30 AAAACATATTATGCTAATGATACTGAAATAGCTAAAGCTGCAGTTTTAGATAAAGATTCA 1560  
-----+-----+-----+-----+-----  
501 K T Y Y A N D T E I A K A A V L D K D S  
520  
1561  
35 GGTGCTTCAAGTAGTGATGCAGGAGTATTTAATTTCTATGTAGCTAAAGATGGATCTACA 1620  
-----+-----+-----+-----+-----  
521 G A S S S D A G V F N F Y V A K D G S T  
540  
1621  
40 AAAGAAGATCAATTAGTTGATGCATTAGCAGTAGGAGCTGTTGCTGGATATAAACTTGCT 1680  
-----+-----+-----+-----+-----  
541 K E D Q L V D A L A V G A V A G Y K L A  
560



```

1681
CCAGTTGTATTAGCTACTGATTCTTTATCTTCTGATCAATCGGTTGCTATAAGCAAAGTT 1740
-----+-----+-----+-----+-----+-----+-----+
5      561  P  V  V  L  A  T  D  S  L  S  S  D  Q  S  V  A  I  S  K  V
580
1741
GTAGGAGAAAAATATTCTAAAGATTTAACACAAGTTGGTCAAGGAATAGCTAATTCAGTT 1800
-----+-----+-----+-----+-----+-----+-----+
10     581  V  G  E  K  Y  S  K  D  L  T  Q  V  G  Q  G  I  A  N  S  V
600
1801  ATAAACAAAATGAAAGATTTATTAGATATG 1830
-----+-----+-----+
15     601  I  N  K  M  K  D  L  L  D  M      610
```

## Appendix 3

5 SEQ ID No. 5. Nucleotide sequence of *slpA* from *Clostridium difficile* strain 170324, PCR type 12, with translation. The putative secretory signal cleavage site ( $\Delta$ ) and site of cleavage to form the two mature SLPs ( $\blacklozenge$ ) are indicated.

```
1
ATGAATAAGAAAAATATAGCAATAGCTATGTCAGGTTTAAACAGTTTCTAGCTTCGGCTGCT
10 60
-----+-----+-----+-----+-----+-----
1 M N K K N I A I A M S G L T V L A S A A
20
61
15 CCTGTTTTTGCTGCAACTACTGGAACACAAGGTTATACTGTAGTTAAAAACGACTGGAAA 120
-----+-----+-----+-----+-----+-----
21 P V F A A T T G T Q G Y T V V K N D W K
40
          Δ
20 121
AAAGCAGTAAACAATTACAAGATGGACTAAAAGATAATAGTATAGGAAAGATAACTGTA 180
-----+-----+-----+-----+-----+-----
41 K A V K Q L Q D G L K D N S I G K I T V
60
25 181
TCTTTTAATGATGGGGTTGTGGGTGAAGTAGCTCCTAAAAGTGCTAATAAGAAAGCGGAC 240
-----+-----+-----+-----+-----+-----
61 S F N D G V V G E V A P K S A N K K A D
80
30 241
AGAGATGCTGCAGCTGAGAAGTTATATAATCTTGTTAACTCAATTAGATAAAATTAGGT 300
-----+-----+-----+-----+-----+-----
81 R D A A A E K L Y N L V N T Q L D K L G
100
35 301
GATGGAGATTATGTTGATTTTTCTGTAGATTATAATTTAGAAAACAAAATAATAACTAAT 360
-----+-----+-----+-----+-----+-----
101 D G D Y V D F S V D Y N L E N K I I T N
120
40 361
CAAGCAGATGCAGAAGCAATTGTTACAAAGTTAAATTCACCTTAATGAGAAAACCTCTTATT 420
-----+-----+-----+-----+-----+-----
121 Q A D A E A I V T K L N S L N E K T L I
140
45 421
GATATAGCAACTAAAGATACTTTTGAATGGTTAGTAAACACAAGATAGTGAAGGTAAA 480
-----+-----+-----+-----+-----+-----
141 D I A T K D T F G M V S K T Q D S E G K
160
```

481  
AATGTTGCTGCAACAAAGGCACTTAAAGTTAAAGATGTTGCTACATTTGGTTTGAAGTCT 540  
-----+-----+-----+-----+-----+-----  
5 161 N V A A T K A L K V K D V A T F G L K S  
180  
541  
GGTGGAAGCGAAGATACTGGATATGTTGTTGAAATGAAAGCAGGAGCTGTAGAGGATAAG 600  
-----+-----+-----+-----+-----+-----  
10 181 G G S E D T G Y V V E M K A G A V E D K  
200  
601  
TATGGTAAAGTTGGAGATAGTACGGCAGGTATTGCAATAAATCTTCCTAGTACTGGACTT 660  
-----+-----+-----+-----+-----+-----  
15 201 Y G K V G D S T A G I A I N L P S T G L  
220  
661  
GAATATGCAGGTAAAGGAACAACAATTGATTTTAATAAAACTTTAAAAGTTGATGTAACA 720  
-----+-----+-----+-----+-----+-----  
20 221 E Y A G K G T T I D F N K T L K V D V T  
240  
721  
GGTGGTTCAACACCTAGTGCTGTAGCTGTAAGTGGTTTTGTAACATAAGATGATACTGAT 780  
-----+-----+-----+-----+-----+-----  
25 241 G G S T P S A V A V S G F V T K D D T D  
260  
781  
TTAGCAAAATCAGGTACTATAAATGTAAGAGTTATAAATGCAAAAGAAGAATCAATTGAT 840  
-----+-----+-----+-----+-----+-----  
30 261 L A K S G T I N V R V I N A K E E S I D  
280  
841  
ATAGATGCAAGCTCATATACATCAGCTGAAAATTTAGCTAAAAGATATGTATTTGATCCA 900  
-----+-----+-----+-----+-----+-----  
35 281 I D A S S Y T S A E N L A K R Y V F D P  
300  
901  
GATGAAATTTCTGAAGCATATAAGGCAATAGTAGCATTACAAAATGATGGTATAGAGTCT 960  
-----+-----+-----+-----+-----+-----  
40 301 D E I S E A Y K A I V A L Q N D G I E S  
320  
961  
AACTTAGTTTCAGTTAGTTAATGGAAAATATCAAGTGATTTTTTATCCAGAAGGTAAAAGA 1020  
-----+-----+-----+-----+-----+-----  
45 321 N L V Q L V N G K Y Q V I F Y P E G K R  
340  
1021  
TTAGAAACTAAATCAGCAAATGATACAATAGCTAGTCAAGATACACCAGCTAAAGTAGTT 1080  
-----+-----+-----+-----+-----+-----  
50 341 L E T K S A N D T I A S Q D T P A K V V  
360  
◆  
1081  
55 ATAAAAGCTAATAAATTAAAAGATTTAAAAGATTATGTAGATGATTTAAAAACATATAAT 1140  
-----+-----+-----+-----+-----+-----

361 I K A N K L K D L K D Y V D D L K T Y N  
380  
1141  
AATACTTATTCAAATGTTGTAACAGTAGCAGGAGAAGATAGAATAGAACTGCTATAGAA 1200  
5  
-----+-----+-----+-----+-----+-----+-----  
381 N T Y S N V V T V A G E D R I E T A I E  
400  
1201  
TTAAGTAGTAAATATTATAATTCTGATGATAAAAAATGCAATAACTGATAAAGCAGTTAAT 1260  
10  
-----+-----+-----+-----+-----+-----+-----  
401 L S S K Y Y N S D D K N A I T D K A V N  
420  
1261  
GATATAGTATTAGTTGGATCTACATCTATAGTTGATGGTCTTGTTCATCACCATTAGCT 1320  
15  
-----+-----+-----+-----+-----+-----+-----  
421 D I V L V G S T S I V D G L V A S P L A  
440  
1321  
TCAGAAAAACAGCTCCATTATTATTAACCTTCAAAAGATAAATTAGATTCATCAGTAAAA 1380  
20  
-----+-----+-----+-----+-----+-----+-----  
441 S E K T A P L L L T S K D K L D S S V K  
460  
1381  
TCTGAAATAAAGAGAGTTATGAACTTAAAGAGTGACACTGGTATAAATACTTCTAAAAAA 1440  
25  
-----+-----+-----+-----+-----+-----+-----  
461 S E I K R V M N L K S D T G I N T S K K  
480  
1441  
GTTTATTTAGCTGGTGGAGTTAATTCTATATCTAAAGATGTAGAAAATGAATTGAAAAAC 1500  
30  
-----+-----+-----+-----+-----+-----+-----  
481 V Y L A G G V N S I S K D V E N E L K N  
500  
1501  
ATGGGTCTTAAAGTTACTAGATTATCAGGAGAAGACAGATACGAAACTTCTTTAGCAATA 1560  
35  
-----+-----+-----+-----+-----+-----+-----  
501 M G L K V T R L S G E D R Y E T S L A I  
520  
1561  
GCTGATGAAATAGGTCTTGATAATGATAAAGCATTGTAGTTGGTGGTACTGGATTAGCA 1620  
40  
-----+-----+-----+-----+-----+-----+-----  
521 A D E I G L D N D K A F V V G G T G L A  
540  
1621  
GATGCTATGAGTATAGCTCCAGTTGCTTCTCAACTTAAAGATGGAGATGCTACTCCAATA 1680  
45  
-----+-----+-----+-----+-----+-----+-----  
541 D A M S I A P V A S Q L K D G D A T P I  
560  
1681  
GTAGTTGTAGATGGAAAAGCAAAAGAAATAAGTGATGATGCTAAGAGTTTCTTAGGAACT 1740  
50  
-----+-----+-----+-----+-----+-----+-----  
561 V V V D G K A K E I S D D A K S F L G T  
580  
1741  
TCTGATGTTGATATAATAGGTGGAAAAAATAGCGTATCTAAAGAGATTGAAGAGTCAATA 1800  
55  
-----+-----+-----+-----+-----+-----+-----

581 S D V D I I G G K N S V S K E I E E S I  
600  
1801  
5 GATAGTGCAACTGGAAAACTCCAGATAGAATAAGTGGAGATGATAGACAAGCAACTAAT 1860  
-----+-----+-----+-----+-----  
601 D S A T G K T P D R I S G D D R Q A T N  
620  
1861  
10 GCTGAAGTTTTAAAAGAAGATGATTATTTACAGATGGTGAAGTTGTGAATTACTTTGTT 1920  
-----+-----+-----+-----+-----  
621 A E V L K E D D Y F T D G E V V N Y F V  
640  
1921  
15 GCAAAAGATGGTTCTACTAAAGAAGATCAATTAGTAGATGCCTTAGCAGCAGCACCAATA 1980  
-----+-----+-----+-----+-----  
641 A K D G S T K E D Q L V D A L A A A P I  
660  
1981  
20 GCAGGTAGATTTAAGGAGTCTCCAGCTCCAATCATACTAGCTACTGATACTTTATCTTCT 2040  
-----+-----+-----+-----+-----  
661 A G R F K E S P A P I I L A T D T L S S  
680  
2041  
25 GACCAAAATGTAGCTGTAAGTAAAGCAGTTCCTAAAGATGGTGGAACCTAAGTTCAA 2100  
-----+-----+-----+-----+-----  
681 D Q N V A V S K A V P K D G G T N L V Q  
700  
2101 GTAGGTAAAGGTATAGCTTCTTCAGTTATAAACAAAATGAAAGATTTATTAGATATG  
2157  
30 -----+-----+-----+-----+-----  
701 V G K G I A S S V I N K M K D L L D M  
719

35

## Appendix 4

5 SEQ ID No 6. Nucleotide sequence of *slpA* from *Clostridium*  
difficile strain 171448, PCR type 12, with translation. The  
putative secretory signal cleavage site ( $\Delta$ ) and site of cleavage to  
form the two mature SLPs ( $\blacklozenge$ ) are indicated.

```
1
10 ATGAATAAGAAAAATATAGCAATAGCTATGTCAGGTTTAAACAGTTTATAGCTTCGGCTGCT 60
    -----+-----+-----+-----+-----+-----+-----+-----
    1 M N K K N I A I A M S G L T V L A S A A
20
    61
15 CCTGTTTTTGGCTGCAACTACTGGAACACAAGGTTATACTGTAGTTAAAAACGACTGGAAA 120
    -----+-----+-----+-----+-----+-----+-----+-----
    21 P V F A A T T G T Q G Y T V V K N D W K
40
    Δ
    121
20 AAAGCAGTAAACAATTACAAGATGGACTAAAAGATAATAGTATAGGAAAGATAACTGTA 180
    -----+-----+-----+-----+-----+-----+-----+-----
    41 K A V K Q L Q D G L K D N S I G K I T V
60
    181
25 TCTTTTAAATGATGGGGTTGTGGGTGAAGTAGCTCCTAAAAGTGCTAATAAGAAAGCGGAC 240
    -----+-----+-----+-----+-----+-----+-----+-----
    61 S F N D G V V G E V A P K S A N K K A D
80
    241
30 AGAGATGCTGCAGCTGAGAAGTTATATAATCTTGTTAACTCAATTAGATAAAATTAGGT 300
    -----+-----+-----+-----+-----+-----+-----+-----
    81 R D A A A E K L Y N L V N T Q L D K L G
100
    301
35 GATGGAGATTATGTTGATTTTTCTGTAGATTATAATTTAGAAAACAAAATAATAACTAAT 360
    -----+-----+-----+-----+-----+-----+-----+-----
    101 D G D Y V D F S V D Y N L E N K I I T N
120
    361
40 CAAGCAGATGCAGAAGCAATTGTTACAAAGTTAAATTCACTTAATGAGAAAACTCTTATT 420
    -----+-----+-----+-----+-----+-----+-----+-----
    121 Q A D A E A I V T K L N S L N E K T L I
140
    421
45 GATATAGCAACTAAAGATACTTTTGGAAATGGTTAGTAAAACACAAGATAGTGGAGGTAAA 480
    -----+-----+-----+-----+-----+-----+-----+-----
    141 D I A T K D T F G M V S K T Q D S G G K
160
```

481  
AATGTTGCTGCAACAAAGGCACTTAAAGTTAAAGATGTTGCTACATTTGGTTTGAAGTCT 540  
-----+-----+-----+-----+-----  
5 161 N V A A T K A L K V K D V A T F G L K S  
180  
541  
GGTGAAGCGAAGATACTGGATATGTTGTTGAAATGAAAGCAGGAGCTGTAGAGGATAAG 600  
-----+-----+-----+-----+-----  
10 181 G G S E D T G Y V V E M K A G A V E D K  
200  
601  
TATGGTAAAGTTGGAGATAGTACGGCAGGTATTGCAATAAATCTTCCTAGTACTGGACTT 660  
-----+-----+-----+-----+-----  
15 201 Y G K V G D S T A G I A I N L P S T G L  
220  
661  
GAATATGCAGGTAAAGGAACAACAATTGATTTTAATAAACTTTAAAAGTTGATGTAACA 720  
-----+-----+-----+-----+-----  
20 221 E Y A G K G T T I D F N K T L K V D V T  
240  
721  
GGTGGTTCAACACCTAGTGCTGTAGCTGTAAGTGGTTTTGTAAGTAAAGATGATACTGAT 780  
-----+-----+-----+-----+-----  
25 241 G G S T P S A V A V S G F V T K D D T D  
260  
781  
TTAGCAAAATCAGGTACTATAAATGTAAGAGTTATAAATGCAAAAGAAGAATCAATTGAT 840  
-----+-----+-----+-----+-----  
30 261 L A K S G T I N V R V I N A K E E S I D  
280  
841  
ATAGATGCAAGCTCATATACATCAGCTGAAAATTTAGCTAAAAGATATGTATTTGATCCA 900  
-----+-----+-----+-----+-----  
35 281 I D A S S Y T S A E N L A K R Y V F D P  
300  
901  
GATGAAATTTCTGAAGCATATAAGGCAATAGTAGCATTACAAAATGATGGTATAGAGTCT 960  
-----+-----+-----+-----+-----  
40 301 D E I S E A Y K A I V A L Q N D G I E S  
320  
961  
AATTTAGTTCAGTTAGTTAATGGAAAATATCAAGTGATTTTTTATCCAGAAGGTAAAAGA 1020  
-----+-----+-----+-----+-----  
45 321 N L V Q L V N G K Y Q V I F Y P E G K R  
340  
1021  
TTAGAAACTAAATCAGCAAATGATACAATAGCTAGTCAAGATACACCAGCTAAAGTAGTT 1080  
-----+-----+-----+-----+-----  
50 341 L E T K S A N D T I A S Q D T P A K V V  
360  
◆  
1081  
55 ATAAAAGCTAATAAATTAAAAGATTTAAAAGATTATGTAGATGATTTAAAAACATATAAT 1140  
-----+-----+-----+-----+-----

361 I K A N K L K D L K D Y V D D L K T Y N  
380  
1141  
AATACTTATTCAAATGTTGTAACAGTAGCAGGAGAAGATAGAATAGAACTGCTATAGAA 1200  
5  
381 N T Y S N V V T V A G E D R I E T A I E  
400  
1201  
TTAAGTAGTAAATATTATAATTCTGATGATAAAAAATGCAATAACTGATAAAGCAGTTAAT 1260  
10  
401 L S S K Y Y N S D D K N A I T D K A V N  
420  
1261  
GATATAGTATTAGTTGGATCTACATCTATAGTTGATGGTCTTGTTGCATCACCATTAGCT 1320  
15  
421 D I V L V G S T S I V D G L V A S P L A  
440  
1321  
TCAGAAAAACAGCTCCATTATTATTAGCTTCAAAAGATAAATTAGATTCATCAGTAAAA 1380  
20  
441 S E K T A P L L L A S K D K L D S S V K  
460  
1381  
TCTGAAATAAAGAGAGTTATGAACTTAAAGAGTGACACTGGTATAAATACTTCTAAAAAA 1440  
25  
461 S E I K R V M N L K S D T G I N T S K K  
480  
1441  
GTTTATTTAGCTGGTGGAGTTAATTCTATATCTAAAGATGTAGAAAATGAATTGAAAAAC 1500  
30  
481 V Y L A G G V N S I S K D V E N E L K N  
500  
1501  
ATGGGTCCTTAAAGTTACTAGATTATCAGGAGAAGACAGATACGAACTTCTTTAGCAATA 1560  
35  
501 M G L K V T R L S G E D R Y E T S L A I  
520  
1561  
GCTGATGAAATAGGTCTTGATAATGATAAAGCATTTGTAGTTGGTGGTACTGGATTAGCA 1620  
40  
521 A D E I G L D N D K A F V V G G T G L A  
540  
1621  
GATGCTATGAGTATAGCTCCAGTTGCTTCTCAACTTAAAGATGGAGATGCTACTCCAATA 1680  
45  
541 D A M S I A P V A S Q L K D G D A T P I  
560  
1681  
GTAGTTGTAGATGGAAAAGCAAAAGAAATAAGTGATGATGCTAAGAGTTTCTTAGGAACT 1740  
50  
561 V V V D G K A K E I S D D A K S F L G T  
580  
1741  
TCTGATGTTGATATAATAGGTGGAAAAAATAGCGTATCTAAAGAGATTGAAGAGTCAATA 1800  
55



581 S D V D I I G G K N S V S K E I E E S I  
600  
1801  
GATAGTGCAACTGGAAAACTCCAGATAGAATAAGTGGAGATGATAGACAAGCAACTAAT 1860  
5  
-----+-----+-----+-----+-----+-----  
601 D S A T G K T P D R I S G D D R Q A T N  
620  
1861  
GCTGAAGTTTTTAAAAGAAGATGATTATTTTACAGATGGTGAAGTTGTGAATTACTTTGTT 1920  
10  
-----+-----+-----+-----+-----+-----  
621 A E V L K E D D Y F T D G E V V N Y F V  
640  
1921  
GCAAAAAGATGGTTCTACTAAAGAAGATCAATTAGTAGATGCCTTAGCAGCAGCACCAATA 1980  
15  
-----+-----+-----+-----+-----+-----  
641 A K D G S T K E D Q L V D A L A A A P I  
660  
1981  
GCAGGTAGATTTAAGGAGTCTCCAGCTCCAATCATACTAGCTACTGATACTTTATCTTCT 2040  
20  
-----+-----+-----+-----+-----+-----  
661 A G R F K E S P A P I I L A T D T L S S  
680  
2041  
GACCAAAATGTAGCTGTAAGTAAAGCAGTTCCTAAAGATGGTGGAACTAACTTAGTTCAA 2100  
25  
-----+-----+-----+-----+-----+-----  
681 D Q N V A V S K A V P K D G G T N L V Q  
700  
2101 GTAGGTAAAGGTATAGCTTCTTCAGTTATAAACAAAATGAAAGATTTATTAGATATG  
2157  
30  
-----+-----+-----+-----+-----+-----  
701 V G K G I A S S V I N K M K D L L D M  
719  
  
35

## Appendix 5

5 SEQ ID No. 7. Nucleotide sequence of *slpA* from *Clostridium difficile* strain 171862, PCR type 17, with translation. The putative secretory signal cleavage site ( $\Delta$ ) and site of cleavage to form the two mature SLPs ( $\blacklozenge$ ) are indicated.

```

1
10 ATGAATAAGAAAACTTAGCAATGGCTATGGCAGCAGTTACTGTTGTGGGTTCTGCAGCG 60
    -----+-----+-----+-----+-----
    1  M  N  K  K  N  L  A  M  A  M  A  A  V  T  V  V  G  S  A  A
20
    61
15 CCAATATTTGCAGATAGTACTACGCCAGGTTATACTGTAGTGAAAAATGATTGGAAAAAA 120
    -----+-----+-----+-----+-----
    21  P  I  F  A  D  S  T  T  P  G  Y  T  V  V  K  N  D  W  K  K
40
        Δ
    121
20 GCAGTAAACAATTACAAGATGGGTTGAAAAATAAACTATATCAACAATAAAGGTGTCT 180
    -----+-----+-----+-----+-----
    41  A  V  K  Q  L  Q  D  G  L  K  N  K  T  I  S  T  I  K  V  S
60
    181
25 TTTAATGGAACTCTGTTGGAGAAGTTACACCAGCCAGTTCTGGAGCAAAAAAAGCAGAT 240
    -----+-----+-----+-----+-----
    61  F  N  G  N  S  V  G  E  V  T  P  A  S  S  G  A  K  K  A  D
80
    241
30 AGAGATGCTGCAGCTGAAAAGTTATATAATTTAGTAAATACACAATTAGATAAACTAGGT 300
    -----+-----+-----+-----+-----
    81  R  D  A  A  A  E  K  L  Y  N  L  V  N  T  Q  L  D  K  L  G
100
    301
35 GATGGAGATTACGTTGACTTTGAAGTAACTTATAATTTAGCTACTCAAATAATTACAAAA 360
    -----+-----+-----+-----+-----
    101 D  G  D  Y  V  D  F  E  V  T  Y  N  L  A  T  Q  I  I  T  K
120
    361
40 GCAGAAGCAGAGGCAGTTCTTACAAAATTACAACAATATAATGATAAAGTACTTATAAAT 420
    -----+-----+-----+-----+-----
    121 A  E  A  E  A  V  L  T  K  L  Q  Q  Y  N  D  K  V  L  I  N
140
    421
45 TCTGCAACAGATACAGTAAAAGGTATGGTATCTGATACACAAGTTGATAGCAAAAATGTT 480
    -----+-----+-----+-----+-----
    141 S  A  T  D  T  V  K  G  M  V  S  D  T  Q  V  D  S  K  N  V
160

```

481  
GCAGCTAACCCACTTAAAGTTAGTGATATGTATACAATACCATCTGCTATTACTGGAAGT 540  
-----+-----+-----+-----+-----+-----+-----  
5 161 A A N P L K V S D M Y T I P S A I T G S  
180  
541  
GATGATTCTGGGTATAGTATTGCTAAACCAACAGAAAAGACTACAaGTTTATTGTATGGT 600  
-----+-----+-----+-----+-----+-----+-----  
10 181 D D S G Y S I A K P T E K T T S L L Y G  
200  
601  
ACGGTTGGTGATGCAACTGCAGGTAAAGCAATAACAGTAGATACAGCTTCAAATGAAGCT 660  
-----+-----+-----+-----+-----+-----+-----  
15 201 T V G D A T A G K A I T V D T A S N E A  
220  
661  
TTTGCTGGAAATGGAAAGGTTATTGACTACAATAAATCATTCAAAGCAACTGTACAAGGA 720  
-----+-----+-----+-----+-----+-----+-----  
20 221 F A G N G K V I D Y N K S F K A T V Q G  
240  
721  
GATGGAACAGTTAAGACAAGCGGGGTTGTACTTAAAGATGCAAGTGATATGGCTGCAACA 780  
-----+-----+-----+-----+-----+-----+-----  
25 241 D G T V K T S G V V L K D A S D M A A T  
260  
781  
GGTACTATAAAAGTTAGAGTTACAAGTGCAAAAGAAGAATCTATTGATGTGGATTCAAGT 840  
-----+-----+-----+-----+-----+-----+-----  
30 261 G T I K V R V T S A K E E S I D V D S S  
280  
841  
TCATATATTAGTGCTGAAAATTTAGCTAAAAAATATGTATTTAATCCTAAAGAGGTTTCT 900  
-----+-----+-----+-----+-----+-----+-----  
35 281 S Y I S A E N L A K K Y V F N P K E V S  
300  
901  
GAAGCTTATAATGCAATAGTTGCATTACAAAATGATGGAATAGAATCTGATTTAGTACAA 960  
-----+-----+-----+-----+-----+-----+-----  
40 301 E A Y N A I V A L Q N D G I E S D L V Q  
320  
961  
TTAGTTAATGGAAAATATCAAGTTATTTTCTATCCAGAAGGAAAAAGATTAGAACTAAA 1020  
-----+-----+-----+-----+-----+-----+-----  
45 321 L V N G K Y Q V I F Y P E G K R L E T K  
340

1021 TCTGCAGATATAATAGCTGATGCAGATAGTCCAGCTAAAATAACTATAAAAGCTAATAAA  
1080  
-----+-----+-----+-----+-----+-----+-----  
341 S A D I I A D A D S P A K I T I K A N K  
5 360  
◆  
1081  
TTAAAAGATTAAAAGATTATGTAGATGATTTAAAACATACAATAATACTTACTCAAAT 1140  
-----+-----+-----+-----+-----+-----+-----  
10 361 L K D L K D Y V D D L K T Y N N T Y S N  
380  
1141  
GTTGTAACAGTAGCAGGAGAAGATAGAATAGAACTGCTATAGAATTAAGTAGTAAATAT 1200  
-----+-----+-----+-----+-----+-----+-----  
15 381 V V T V A G E D R I E T A I E L S S K Y  
400  
1201  
TATAATTCTGATGATAAAAATGCAATAACTGATGATGCAGTTAATAATATAGTATTAGTT 1260  
-----+-----+-----+-----+-----+-----+-----  
20 401 Y N S D D K N A I T D D A V N N I V L V  
420  
1261  
GGATCTACATCTATAGTTGATGGTCTTGTTCATCACCATTAGCTTCAGAAAAACAGCT 1320  
-----+-----+-----+-----+-----+-----+-----  
25 421 G S T S I V D G L V A S P L A S E K T A  
440  
1321  
CCATTATTATTAACCTTCAAAAGATAAATTAGATTCATCAGTAAAATCTGAGATAAAAAGA 1380  
-----+-----+-----+-----+-----+-----+-----  
30 441 P L L L T S K D K L D S S V K S E I K R  
460  
1381  
GTTATGAACCTAAAGAGTGATACTGGTATAAATACTTCTAAAAAAGTTTATTTAGCTGGT 1440  
-----+-----+-----+-----+-----+-----+-----  
35 461 V M N L K S D T G I N T S K K V Y L A G  
480  
1441  
GGAGTTAATTCTATATCTAAAGATGTAGAAGATGAATTGAAAAATATGGGCCTTAAAGTT 1500  
-----+-----+-----+-----+-----+-----+-----  
40 481 G V N S I S K D V E D E L K N M G L K V  
500  
1501  
ACTAGATTATCAGGAGAAGACAGATACGAACTTCTTTAGCAATAGCTGATGAAATAGGT 1560  
-----+-----+-----+-----+-----+-----+-----  
45 501 T R L S G E D R Y E T S L A I A D E I G  
520  
1561  
CTTGATAATGATAAAGCATTGTAGTTGGTGGTACTGGATTGGCAGATGCTATGAGTATA 1620  
-----+-----+-----+-----+-----+-----+-----  
50 521 L D N D K A F V V G G T G L A D A M S I  
540  
1621  
55 GCTCCAGTTGCTTCTCAACTTAAAGATGGAGATGCTACTCCAATAGTAGTTGTAGATGGA 1680  
-----+-----+-----+-----+-----+-----+-----

541 A P V A S Q L K D G D A T P I V V V D G  
560  
1681  
AAAGCAAAGAAATAAGTGATGATGCTAAGAGTTTCTTAGGAACTTCTGATGTTGATATA 1740  
5  
-----+-----+-----+-----+-----+-----  
561 K A K E I S D D A K S F L G T S D V D I  
580  
1741  
ATAGGTGGAAAAAATAGCGTATCTAAAGAGATTGAAGAGTCAATAGATAGTGCAACTGGA 1800  
10  
-----+-----+-----+-----+-----+-----  
581 I G G K N S V S K E I E E S I D S A T G  
600  
1801  
AAAACCTCCAGATAGAATAAGTGAGATGACAGACAAGCAACTAATGCTGAAGTTTTAAAA 1860  
15  
-----+-----+-----+-----+-----+-----  
601 K T P D R I S G D D R Q A T N A E V L K  
620  
1861  
GAAGATGATTATTTCAAAGATGGTGAAGTTGTGAATTACTTTGTTGCAAAAGATGGTTCT 1920  
20  
-----+-----+-----+-----+-----+-----  
621 E D D Y F K D G E V V N Y F V A K D G S  
640  
1921  
ACTAAAGAAGATCAATTAGTAGATGCATTAGCAGCAGCACCAATAGCAGGTAGATTTAAG 1980  
25  
-----+-----+-----+-----+-----+-----  
641 T K E D Q L V D A L A A A P I A G R F K  
660  
1981  
GAGTCTCCAGCTCCAATCATACTAGCTACTGATACTTTATCTTCTGACCAAAATGTAGCT 2040  
30  
-----+-----+-----+-----+-----+-----  
661 E S P A P I I L A T D T L S S D Q N V A  
680  
2041  
GTAAGTAAAGCAGTTCCTAAAGATGGTGGAACCTAAGTTCAAGTAGGTAAAGGTATA 2100  
35  
-----+-----+-----+-----+-----+-----  
681 V S K A V P K D G G T N L V Q V G K G I  
700  
2101 GCTTCTTCAGTTATAAACAAAATGAAAGATTTATTAGATATGTAA 2145  
40  
-----+-----+-----+-----+-----+-----  
701 A S S V I N K M K D L L D M \* 715

## Appendix 6

5 SEQ ID No 8. Nucleotide sequence of *slpA* from *Clostridium*  
*difficile* strain 173644, PCR type 31, with translation. The  
putative secretory signal cleavage site (Δ) and site of cleavage to  
form the two mature SLPs (◆) are indicated.

10  
15  
20  
25  
30  
35  
40  
45  
50  
55

1  
61  
121  
181  
241  
301  
361  
421  
481

ATGAATAAGAAGGATATAGCAATAGCTATGTCAGGATTAACAGTATTAGCTTCTGCAGCA  
CCTGTATTTGCTGCTAGTAGTTTTACAGCAGATTATAATTATACTGTAGTGCAAGGAAAA  
TATCAAAAAGTTATAACTGGATTACAAGATGGTTTAAAAAATGGAAAAATAACAAATATT  
GATGTAATATTTGATGGAAGTTCAATTGGTGAGGTAGTGCCAGGTTCTGATGCTGCAGCT  
GCAGCTACTAAATTA AAAAGTTTAGTTGATGATAAGTTAGATAACTTAGGTGATGGAAAA  
TACGTTCAATTTAATGTTACTTTATACTACTAAATCTATAATAACTAAAAGCAGAATTAAAA  
AATTATTATAATCAATTAGAAAAGTAGTAAAGATAGAATACTTATAGGAAAATGAACCTCAA  
GATACAGGA ACTAAAGGTCTTATAAAAAGCTGATACTGATGGTACTACTGCTGTTGCAGCA  
GCTGCACCATTTGAAAATTATCAGATATATTTACGTTTAGTTATGATGAAGTAACAGGTGTA

60  
120  
180  
240  
300  
360  
420  
480  
540

1  
21  
41  
61  
81  
101  
121  
141  
161

M N K K D I A I A M S G L T V L A S A A  
P V F A A S S F T A D Y N Y T V V Q G K  
Y Q K V I T G L Q D G L K N G K I T N I  
D V I F D G S S I G E V V P G S D A A A  
A A T K L K S L V D D K L D N L G D G K  
Y V Q F N V T Y T T K S I I T K A E L K  
N Y Y N Q L E S S K D R I L I G N E P Q  
D T G T K G L I K A D T D G T T A V A A  
A A P L K L S D I F T F S Y D E V T G V

541  
CTTAAAGCAGAACCAACAAGTAAAGTAAGCGCTGGTAAAGTTCAAGGTCTAAAAATATGGA 600  
-----+-----+-----+-----+-----  
181 L K A E P T S K V S A G K V Q G L K Y G  
200  
601  
AATACAGGAGCAACTAACTATACTTCTGGAGCTGAAATATCTGTTCTACTACAGGCTTA 660  
-----+-----+-----+-----+-----  
201 N T G A T N Y T S G A E I S V P T T G L  
220  
661  
ACATTAAGTCTGATACAACTGCAACAACAGATGTAAATATTTCTGATGTTATGAGTGCA 720  
-----+-----+-----+-----+-----  
221 T L T A D T T A T T D V N I S D V M S A  
240  
721  
TTTAAATTTAATGGTACTGATACGATTAGTGGATTCCCAGCTGGTTCATCAGCTTCTACT 780  
-----+-----+-----+-----+-----  
241 F K F N G T D T I S G F P A G S S A S T  
260  
781  
CTTAGAGCAAGTATAAAAGTAATAAATGCAAAAGAAGAATCTATAGATGTTGATTCAAGT 840  
-----+-----+-----+-----+-----  
261 L R A S I K V I N A K E E S I D V D S S  
280  
841  
TCACATAGAACAGCTGAAGATTTAGCTGAAAAATATGTATTTAAACCAGAAGATGTGAAT 900  
-----+-----+-----+-----+-----  
281 S H R T A E D L A E K Y V F K P E D V N  
300  
901  
AAAAGTATGAGGCACTGACTGATTTATATAAAGAAGGTATAACAAGTAATCTTATCACT 960  
-----+-----+-----+-----+-----  
301 K T Y E A L T D L Y K E G I T S N L I T  
320  
961  
CAAGATGGTGGAAAATATCAAGTTGTTTTATTTGCTCAAGGAAAGAGATTAAGTACTAAA 1020  
-----+-----+-----+-----+-----  
321 Q D G G K Y Q V V L F A Q G K R L T T K  
340  
1021  
GGAGCAACTGGAACCTTAGCAGATGAAAATTCTCCTCTTAAAGTAACAATAAAAGCAGAT 1080  
-----+-----+-----+-----+-----  
341 G A T G T L A D E N S P L K V T I K A D  
360  
◆  
1081  
AAAGTAAAAGACTTAAAAGATTATGTTGAAGATTTAAAAAATGCTAACAATGGATATTCA 1140  
-----+-----+-----+-----+-----  
361 K V K D L K D Y V E D L K N A N N G Y S  
380  
1141  
AATTCTGTTGTTGTAGCAGGTGAAGATAGAATAGAAACAGCAATAGAGTTAAGTAGCAAA 1200  
-----+-----+-----+-----+-----

381 N S V V V A G E D R I E T A I E L S S K  
400  
1201  
TACTATAACTCTGATGATGACAATGCAATAACTAAAGATCCAGTTAACAATGTTGTTTTTA 1260  
5  
401 Y Y N S D D D N A I T K D P V N N V V L  
420  
1261  
GTTGGTTCTCAAGCTGTAGTTGATGGGCTTGTAGCTTCACCTTTAGCATCTGAAAAAAGA 1320  
10  
421 V G S Q A V V D G L V A S P L A S E K R  
440  
1321  
GCTCCTTTACTATTAACTTCAGCAGGAAAATTAGATTCAAGTGTTAAAGCTGAGTTGAAA 1380  
15  
441 A P L L L T S A G K L D S S V K A E L K  
460  
1381  
AGAGTAATGGATTAAAAATCTACAACAGGTGTAAATACTTCTAAAAAAGTTTACTTAGCT 1440  
20  
461 R V M D L K S T T G V N T S K K V Y L A  
480  
1441  
GGTGGAGTAAACTCTATATCTAAAGATGTAGAAAATGAATTAAAAGATATGGGACTTAAA 1500  
25  
481 G G V N S I S K D V E N E L K D M G L K  
500  
1501  
GTTACAAGATTATCAGGAGATGATAGATATGAACTTCTTTAGCTATAGCTGATGAAATA 1560  
30  
501 V T R L S G D D R Y E T S L A I A D E I  
520  
1561  
GGTCTTGATAATGATAAAGCTTTTGTAGTTGGAGGAACAGGATTAGCGGATGCTATGAGT 1620  
35  
521 G L D N D K A F V V G G T G L A D A M S  
540  
1621  
ATAGCTCCAGTTGCTTCTCAATTAAGAACTCAAATGGAGAACTTGACTTAAAAGGTGAT 1680  
40  
541 I A P V A S Q L R N S N G E L D L K G D  
560  
1681  
GCAACTCCAATAGTAGTTGTTGATGGAAAAGCTAAAGATATAAATTCTGAAGTAAAAGAT 1740  
45  
561 A T P I V V V D G K A K D I N S E V K D  
580  
1741  
TTCTTAGATGATTCACAAGTTGATATAATAGGTGGTGTAATAGTGTCTTCTAAAGAAGTA 1800  
50  
581 F L D D S Q V D I I G G V N S V S K E V  
600  
1801  
ATGGAAGCAATAGATGATGCTACTGGAAAATCACCTGAGAGATATAGTGGAAGATAGA 1860  
55



601 M E A I D D A T G K S P E R Y S G E D R  
620  
1861  
5 CAAGCAACAAATGCTAAAGTTATAAAGAAGATGATTTCTTTAAAAATGGAGAAGTTACA 1920  
-----+-----+-----+-----+-----+-----  
621 Q A T N A K V I K E D D F F K N G E V T  
640  
1921  
10 AACTTCTTTGTAGCTAAAGATGGTTCAACTAAAGAAGATCAATTAGTAGATGCTTTAGCA 1980  
-----+-----+-----+-----+-----+-----  
641 N F F V A K D G S T K E D Q L V D A L A  
660  
1981  
15 GGTGCTGCAATTGCTGGTAACTTTGGTGTAAACAGTAGATAATGAAGGAAAACCTACAGTT 2040  
-----+-----+-----+-----+-----+-----  
661 G A A I A G N F G V T V D N E G K P T V  
680  
2041  
20 GCTGATAAAAAAGCTTCTCCAGCACCAATTGTTTTAGCAACAGATTCTTTATCTTCTGAT 2100  
-----+-----+-----+-----+-----+-----  
681 A D K K A S P A P I V L A T D S L S S D  
700  
2101  
25 CAAAATGTAGCTATAAGTAAAGCTGTAAATGATGACGCTAATACTAAGAATCTAGTTCAA 2160  
-----+-----+-----+-----+-----+-----  
701 Q N V A I S K A V N D D A N T K N L V Q  
720  
2161 GTTGGTAAAGGTATAGCTACTTCAGTTGTAAGTAAAATAAAAGATTTATTAGATATG  
2217  
30 -----+-----+-----+-----+-----+-----  
721 V G K G I A T S V V S K I K D L L D M  
739  
  
35

## Appendix 7

5 SEQ ID No 9. Nucleotide sequence of *slpA* from *Clostridium difficile* strain 170444, PCR type 46, with translation. The putative secretory signal cleavage site ( $\Delta$ ) and site of cleavage to form the two mature SLPs ( $\blacklozenge$ ) are indicated.

```

1
10 ATGAATAAGAAAAATATAGCAATAGCTATGTCAGGTTTAACAGTTTTAGCTTCGGCTGCT 60
    -----+-----+-----+-----+-----+-----
    1  M  N  K  K  N  I  A  I  A  M  S  G  L  T  V  L  A  S  A  A
20
    61
15 CCTGTTTTTGCTGCAACTACTGGAACACAAGGTTATACTGTAGTTAAAAACGACTGGAAA 120
    -----+-----+-----+-----+-----+-----
    21  P  V  F  A  A  T  T  G  T  Q  G  Y  T  V  V  K  N  D  W  K
40
    Δ
    121
20 AAAGCAGTAAACAATTACAAGATGGACTAAAAGATAATAGTATAGGAAAGATAACTGTA 180
    -----+-----+-----+-----+-----+-----
    41  K  A  V  K  Q  L  Q  D  G  L  K  D  N  S  I  G  K  I  T  V
60
    181
25 TCTTTTAATGATGGGGTTGTGGGTGAAGTAGCTCCTAAAAGTGCTAATAAGAAAGCGGAC 240
    -----+-----+-----+-----+-----+-----
    61  S  F  N  D  G  V  V  G  E  V  A  P  K  S  A  N  K  K  A  D
80
    241
30 AGAGATGCTGCAGCTGAGAAGTTATATAATCTTGTTAACTCAATTAGATAAATTAGGT 300
    -----+-----+-----+-----+-----+-----
    81  R  D  A  A  A  E  K  L  Y  N  L  V  N  T  Q  L  D  K  L  G
100
    301
35 GATGGAGATTATGTTGATTTTTCTGTAGATTATAATTTAGAAAAAAAAAATAACTAAT 360
    -----+-----+-----+-----+-----+-----
    101 D  G  D  Y  V  D  F  S  V  D  Y  N  L  E  K  K  I  I  T  N
120
    361
40 CAAGCAGATGCAGAAGCAATTGTTACAAAGTTAAATTCACTTAATGAGAAAACCTCTTATT 420
    -----+-----+-----+-----+-----+-----
    121 Q  A  D  A  E  A  I  V  T  K  L  N  S  L  N  E  K  T  L  I
140
    421
45 GATATAGCAACTAAAGATACTTTTGGAATGGTTAGTAAAAACACAAGATAGTGAAGGTAAA 480
    -----+-----+-----+-----+-----+-----
    141 D  I  A  T  K  D  T  F  G  M  V  S  K  T  Q  D  S  E  G  K
160
```

481  
AATGTTGCTGCAACAAAGGCACCTTAAAGTTAAAGATGTTGCTACATTTGGTTTGAAGTCT 540  
-----+-----+-----+-----+-----+-----  
5 161 N V A A T K A L K V K D V A T F G L K S  
180  
541  
GGTGAAGCGAAGATACTGGATATGTTATTGAAATGAAAGCAGGAGCTGTAGAGGATAAG 600  
-----+-----+-----+-----+-----+-----  
10 181 G G S E D T G Y V I E M K A G A V E D K  
200  
601  
TATGGTAAAGTTGGAGATAGTACGGCAGGTATTGCAATAAAATCTTCCTAGTACTGGACTT 660  
-----+-----+-----+-----+-----+-----  
15 201 Y G K V G D S T A G I A I N L P S T G L  
220  
661  
GAATATGCAGGTAAAGGAACAACAATTGATTTTAATAAACTTTAAAAGTTGATGTAACA 720  
-----+-----+-----+-----+-----+-----  
20 221 E Y A G K G T T I D F N K T L K V D V T  
240  
721  
GGTGGTTCAACACCTAGTGCTGTAGCTGTAAGTGGTTTTGTAAGTAAAGATGATACTGAT 780  
-----+-----+-----+-----+-----+-----  
25 241 G G S T P S A V A V S G F V T K D D T D  
260  
781  
TTAGCAAAATCAGGTACTATAAATGTAAGAGTTATAAATGCAAAAGAAGAATCAATTGAT 840  
-----+-----+-----+-----+-----+-----  
30 261 L A K S G T I N V R V I N A K E E S I D  
280  
841  
ATAGATGCAAGCTCATATACATCAGCTGAAAATTTAGCTAAAAGACATGTATTTGATCCA 900  
-----+-----+-----+-----+-----+-----  
35 281 I D A S S Y T S A E N L A K R H V F D P  
300  
901  
GATGAAATTTCTGAAGCATATAAGGCAATAGTAGCATTACAAAATGATGGTATAGAGTCT 960  
-----+-----+-----+-----+-----+-----  
40 301 D E I S E A Y K A I V A L Q N D G I E S  
320  
961  
AATTTAGTTCAGTTAGTTAATGAAAAATATCAAGTGATTTTTTATCCAGAAGGTAAAAGA 1020  
-----+-----+-----+-----+-----+-----  
45 321 N L V Q L V N G K Y Q V I F Y P E G K R  
340

341 L E T K S A N D T I A S Q D T P A K V V  
360

10

361 I K A N K L K D L K D Y V D D L K T Y N  
380

15

381 N T Y S N V V T V A G E D R I E T A I E  
400

20

401 L S S K Y Y N S D D K N A I T D K A V N  
420

25

421 D I V L V G S T S I V D G L V A S P L A  
440

30

441 S E K T A P L L L T S K D K L D S S V K  
460

35

461 S E I K R V M N L K S D T G I N T S K K  
480

40

481 V Y L A G G V N S I S K D V E N E L K N

45

-----+-----+-----+-----+-----  
501 M G L K V T R L S G E D R Y E T S L A I  
520

50

521 A D E I G L D N D K A F V V G G T G L A

55

541 D A M S I A P V A S Q L K D G D A T P I  
560  
1681  
GTAGTTGTAGATGGAAAAGCAAAAGAAATAAGTGATGATGCTAAGAGTTTCTTAGGAACT 1740  
5  
-----+-----+-----+-----+-----+-----  
561 V V V D G K A K E I S D D A K S F L G T  
580  
1741  
TCTGATGTTGATATAATAGGTGGAAAAAATAGCGTATCTAAAGAGATTGAAGAGTCAATA 1800  
10  
-----+-----+-----+-----+-----+-----  
581 S D V D I I G G K N S V S K E I E E S I  
600  
1801  
GATAGTGCAACTGGAAAAACTCCAGATAGAATAAGTGGAGATGATAGACAAGCAACTAAT 1860  
15  
-----+-----+-----+-----+-----+-----  
601 D S A T G K T P D R I S G D D R Q A T N  
620  
1861  
GCTGAAGTTTTTAAAAGAAGATGATTATTTTACAGATGGTGAAGTTGTGAATTACTTTGTT 1920  
20  
-----+-----+-----+-----+-----+-----  
621 A E V L K E D D Y F T D G E V V N Y F V  
640  
1921  
GCAAAAGATGGTTCTACTAAAGAAGATCAATTAGTAGATGCCTTAGCAGCAGCACCAATA 1980  
25  
-----+-----+-----+-----+-----+-----  
641 A K D G S T K E D Q L V D A L A A A P I  
660  
1981  
GCAGGTAGATTTAAGGAGTCTCCAGCTCCAATCATACTAGCTACTGATACTTTATCTTCT 2040  
30  
-----+-----+-----+-----+-----+-----  
661 A G R F K E S P A P I I L A T D T L S S  
680  
2041  
GACCAAATGTAGCTGTAAGTAAAGCAGTTCCTAAAGATGGTGGAACTAACTTAGTTCAA 2100  
35  
-----+-----+-----+-----+-----+-----  
681 D Q N V A V S K A V P K D G G T N L V Q  
700  
2101 GTAGGTAAAGGTATAGCTTCTTCAGTTATAAACAAAATGAAAGATTTATTAGATATG  
2157  
40  
-----+-----+-----+-----+-----+-----  
701 V G K G I A S S V I N K M K D L L D M  
719  
  
45

## Appendix 8

5 SEQ ID No 10. Nucleotide sequence of *slpA* from *Clostridium difficile* strain 170426, PCR type 92, with translation. The putative secretory signal cleavage site ( $\Delta$ ) and site of cleavage to form the two mature SLPs ( $\blacklozenge$ ) are indicated.

```

1
10 ATGAATAAGAAAAATATAGCAATAGCTATGTCAGGTTTAACAGTTTATAGCTTCGGCTGCT 60
    -----+-----+-----+-----+-----+-----+-----+
    1  M  N  K  K  N  I  A  I  A  M  S  G  L  T  V  L  A  S  A  A
20
    61
15 CCTGTTTTTGTCTGCAACTACTGGAACACAAGGTTATACTGTAGTTAAAAACGACTGGAAA 120
    -----+-----+-----+-----+-----+-----+-----+
    21  P  V  F  A  A  T  T  G  T  Q  G  Y  T  V  V  K  N  D  W  K
40
        Δ
    121
20 AAAGCAGTAAACAATTACAGGATGGACTAAAAGATAATAGTATAGGAAAGATAACTGTA 180
    -----+-----+-----+-----+-----+-----+-----+
    41  K  A  V  K  Q  L  Q  D  G  L  K  D  N  S  I  G  K  I  T  V
60
    181
25 TCTTTTAATGATGGGGTTGTGGGTGAAGTAGCTCCTAAAAGTGCTAATAAGAAAGCGGAC 240
    -----+-----+-----+-----+-----+-----+-----+
    61  S  F  N  D  G  V  V  G  E  V  A  P  K  S  A  N  K  K  A  D
80
    241
30 AGAGATGCTGCAGCTGAGAAGTTATATAATCTTGTTAACTCAATTAGATAAATTAGGT 300
    -----+-----+-----+-----+-----+-----+-----+
    81  R  D  A  A  A  E  K  L  Y  N  L  V  N  T  Q  L  D  K  L  G
100
    301
35 GATGGAGATTATGTTGATTTTTCTGTAGATTATAATTTAGAAAAAAAAATAATACTAAT 360
    -----+-----+-----+-----+-----+-----+-----+
    101 D  G  D  Y  V  D  F  S  V  D  Y  N  L  E  K  K  I  I  T  N
120
    361
40 CAAGCAGATGCAGAAGCAATTGTTACAAAGTTAAATTCCTTAATGAGAAAACCTCTTATT 420
    -----+-----+-----+-----+-----+-----+-----+
    121 Q  A  D  A  E  A  I  V  T  K  L  N  S  L  N  E  K  T  L  I
140
    421
45 GATATAGCAACTAAAGATACTTTTGGAAATGGTTAGTAAAACACAAGATAGTGAAGGTAAA 480
    -----+-----+-----+-----+-----+-----+-----+
    +
    141 D  I  A  T  K  D  T  F  G  M  V  S  K  T  Q  D  S  E  G  K
50
    160
    481
55 AATGTTGCTGCAACAAAGGCACTTAAAGTTAAAGATGTTGCTACATTTGGTTTGAAGTCT 540
    -----+-----+-----+-----+-----+-----+-----+
    161 N  V  A  A  T  K  A  L  K  V  K  D  V  A  T  F  G  L  K  S
180

```

541  
GGTGGGAAGCGAAGATACTGGATATGTTGTTGAAATGAAAGCAGGAGCTGTAGAGGATAAG 600  
-----+-----+-----+-----+-----+-----  
181 G G S E D T G Y V V E M K A G A V E D K  
200  
601  
TATGGTAAAGTTGGAGATAGTACGGCAGGTATTGCAATAAATCTTCCTAGTACTGGACTT 660  
-----+-----+-----+-----+-----+-----  
201 Y G K V G D S T A G I A I N L P S T G L  
220  
661  
GAATATGCAGGTAAAGGAACAACAATTGATTTTAATAAAACTTTAAAAGTTGATGTAACA 720  
-----+-----+-----+-----+-----+-----  
221 E Y A G K G T T I D F N K T L K V D V T  
240  
721  
GGTGGTTCAACACCTAGTGCTGTAGCTGTAAGTGGTTTTGTAACATAAGATGATACTGAT 780  
-----+-----+-----+-----+-----+-----  
241 G G S T P S A V A V S G F V T K D D T D  
260  
781  
TTAGCAAAATCAGGTACTATAAATGTAAGAGTTATAAATGCAAAAGAAGAATCAATTGAT 840  
-----+-----+-----+-----+-----+-----  
261 L A K S G T I N V R V I N A K E E S I D  
280  
841  
ATAGATGCAAGCTCATATACATCAGCTGAAAATTTAGCTAAAAGATATGTATTTGATCCA 900  
-----+-----+-----+-----+-----+-----  
281 I D A S S Y T S A E N L A K R Y V F D P  
300  
901  
GATGAAATTTCTGAAGCATATAAGGCAATAGTAGCATTACAAAATGATGGTATAGAGTCT 960  
-----+-----+-----+-----+-----+-----  
301 D E I S E A Y K A I V A L Q N D G I E S  
320  
961  
AATTTAGTTCAGTTAGTTAATGGAAAATATCAAGTGATTTTTTATCCAGAAGGTAAAAGA 1020  
-----+-----+-----+-----+-----+-----  
321 N L V Q L V N G K Y Q V I F Y P E G K R  
340  
1021  
TTAGAACTAAATCAGCAAATGATACAATAGCTAGTCAAGATACACCAGCTAAAGTAGTT 1080  
-----+-----+-----+-----+-----+-----  
341 L E T K S A N D T I A S Q D T P A K V V  
360  
◆  
1081  
ATAAAAGCTAATAAATTTAAAAGATTTAAAAGATTATGTAGATGATTTAAAAACATATAAT 1140  
-----+-----+-----+-----+-----+-----  
361 I K A N K L K D L K D Y V D D L K T Y N  
380  
1141  
AATACTTATTCAAATGTTGTAACAGTAGCAGGAGAAGATAGAATAGAACTGCTATAGAA 1200  
-----+-----+-----+-----+-----+-----





601 D S A T G K T P D R I S G D D R Q A T N  
620  
1861  
5 GCTGAAGTTTAAAAGAAGATGATTATTTACAGATGGTGAAGTTGTGAATTACTTTGTT 1920  
-----+-----+-----+-----+-----+-----  
621 A E V L K E D D Y F T D G E V V N Y F V  
640  
1921  
10 GCAAAAGATGGTTCTACTAAAGAAGATCAATTAGTAGATGCCTTAGCAGCAGCACCAATA 1980  
-----+-----+-----+-----+-----+-----  
641 A K D G S T K E D Q L V D A L A A A P I  
660  
1981  
15 GCAGGTAGATTTAAGGAGTCTCCAGCTCCAATCATACTAGCTACTGATACTTTATCTTCT 2040  
-----+-----+-----+-----+-----+-----  
661 A G R F K E S P A P I I L A T D T L S S  
680  
2041  
20 GACCAAAATGTAGCTGTAAGTAAAGCAGTTCCTAAAGATGGTGGAACCTAAGTTCAA 2100  
-----+-----+-----+-----+-----+-----  
681 D Q N V A V S K A V P K D G G T N L V Q  
700  
2101 GTAGGTAAAGGTATAGCTTCTTCAGTTATAAACAAAATGAAAGATTTATTAGATATG  
2157  
25 -----+-----+-----+-----+-----+-----  
701 V G K G I A S S V I N K M K D L L D M  
719

30

Claims

1. A vaccine for the treatment or prophylaxis of *C. difficile* associated disease, the vaccine comprising a *C. difficile* gene or a *C. difficile* peptide/polypeptide or a derivative or fragment or mutant or variant thereof which is immunogenic in humans.  
5
2. A vaccine for the treatment or prophylaxis of *C. difficile* associated disease, the vaccine comprising a *C. difficile* gene or *C. difficile* peptide/polypeptide or a derivative or fragment or mutant or variant thereof to which immunoreactivity is detected in individuals who have recovered from *C. difficile* infection.  
10
3. A vaccine as claimed in claim 1 or 2 wherein the gene encodes a *C. difficile* surface layer protein, SlpA or variant or homologue thereof.  
15
4. A vaccine as claimed in claim 1 or 2 wherein the peptide/polypeptide is a *C. difficile* surface layer protein, SlpA or variant or homologue thereof.
5. A vaccine as claimed in any of claims 1 to 4 wherein the vaccine comprises a chimeric nucleic acid sequence.  
20
6. A vaccine as claimed in 5 wherein the chimeric nucleic acid sequence is derived from the 5' end of the gene, encoding the mature N-terminal moiety of SlpA from *C. difficile*.  
25
7. A vaccine as claimed in any of claims 1 to 4 wherein the vaccine comprises a chimeric peptide/polypeptide.
8. A vaccine as claimed in 7 wherein the amino acid sequence of the chimeric peptide/polypeptide is derived from the mature N-terminal moiety of SlpA from *C. difficile*.  
30

9. A vaccine as claimed in any of claims 1 to 8 wherein the vaccine contains an amino acid sequence SEQ ID No.1 or a derivative or fragment or mutant or variant thereof.
- 5 10. A vaccine as claimed in any of claims 1 to 8 wherein the vaccine contains an amino acid sequence SEQ ID No.2 or a derivative or fragment or mutant or variant thereof.
- 10 11. A vaccine as claimed in any of claims 1 to 8 wherein the vaccine contains a nucleotide sequence SEQ ID No.3 or a derivative or fragment or mutant or variant thereof.
- 15 12. A vaccine as claimed in any of claims 1 to 8 wherein the vaccine contains a nucleotide sequence SEQ ID No.4 or a derivative or fragment or mutant or variant thereof.
- 20 13. A vaccine as claimed in any of claims 1 to 8 wherein the vaccine contains a nucleotide sequence SEQ ID No.5 or a derivative or fragment or mutant or variant thereof.
- 25 14. A vaccine as claimed in any of claims 1 to 8 wherein the vaccine contains a nucleotide sequence SEQ ID No.6 or a derivative or fragment or mutant or variant thereof.
- 30 15. A vaccine as claimed in any of claims 1 to 8 wherein the vaccine contains a nucleotide sequence SEQ ID No.7 or a derivative or fragment or mutant or variant thereof.
16. A vaccine as claimed in any of claims 1 to 8 wherein the vaccine contains a nucleotide sequence SEQ ID No.8 or a derivative or fragment or mutant or variant thereof.

17. A vaccine as claimed in any of claims 1 to 8 wherein the vaccine contains a nucleotide sequence SEQ ID No.9 or a derivative or fragment or mutant or variant thereof.
- 5 18. A vaccine as claimed in any of claims 1 to 8 wherein the vaccine contains a nucleotide sequence SEQ ID No.10 or a derivative or fragment or mutant or variant thereof.
- 10 19. A vaccine as claimed in any preceding claim in combination with at least one other *C. difficile* sub-unit.
- 15 20. A vaccine for the treatment or prophylaxis of *C. difficile* associated disease, the vaccine comprising the mature N-terminal moiety of a surface layer protein, SlpA of *C. difficile* or variant or homologue thereof which is immunogenic in humans.
21. A vaccine as claimed in claim 20 wherein the N-terminal moiety of SlpA contains an amino acid sequence SEQ ID No. 1.
- 20 22. A vaccine as claimed in claim 20 wherein the N-terminal moiety of SlpA contains an amino acid sequence SEQ ID No. 2.
- 25 23. A vaccine for the treatment or prophylaxis of *C. difficile* associated disease, the vaccine comprising an immunodominant epitope derived from a *C. difficile* gene or a *C. difficile* peptide/polypeptide or a derivative or fragment or mutant or variant thereof which is immunogenic in humans.
24. A vaccine as claimed in any preceding claim comprising a pharmaceutically acceptable carrier.
- 30 25. A vaccine as claimed in any preceding claim in combination with a pharmacologically suitable adjuvant.

26. A vaccine as claimed in claim 25 wherein the adjuvant is interleukin 12.
27. A vaccine as claimed in claim 25 or 26 wherein the adjuvant is a heat shock protein.
- 5 28. A vaccine as claimed in any preceding claim comprising at least one other pharmaceutical product.
- 10 29. A vaccine as claimed in claim 28 wherein the pharmaceutical product is an antibiotic.
30. A vaccine as claimed in claim 29 wherein the antibiotic is selected from one or more metronidazole, amoxycillin, tetracycline or erythromycin, clarithromycin or tinidazole.
- 15 31. A vaccine as claimed in claim 28 wherein the pharmaceutical product comprises an acid-suppressing agent such as omeprazole or bismuth salts.
- 20 32. A vaccine as claimed in any preceding claim in a form for oral administration.
33. A vaccine as claimed in any preceding claim in a form for intranasal administration.
- 25 34. A vaccine as claimed in any preceding claim in a form for intravenous administration.
35. A vaccine as claimed in any preceding claim in a form for intramuscular administration.
- 30 36. A vaccine as claimed in any of claims 1 to 35 including a peptide delivery system.

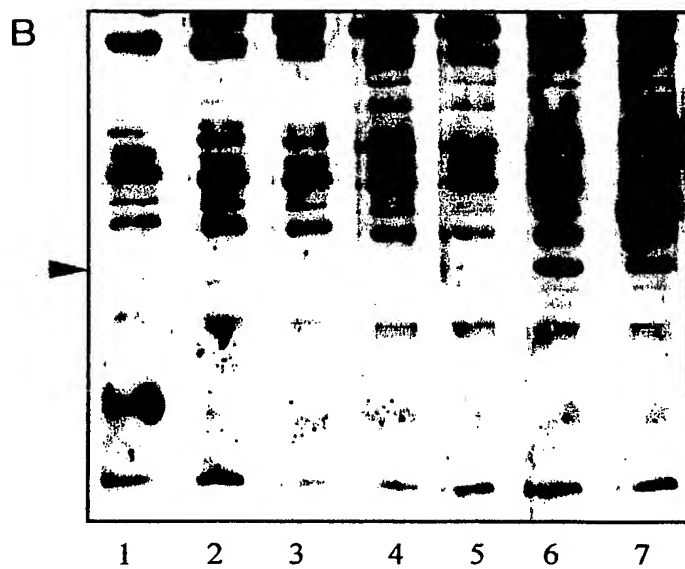
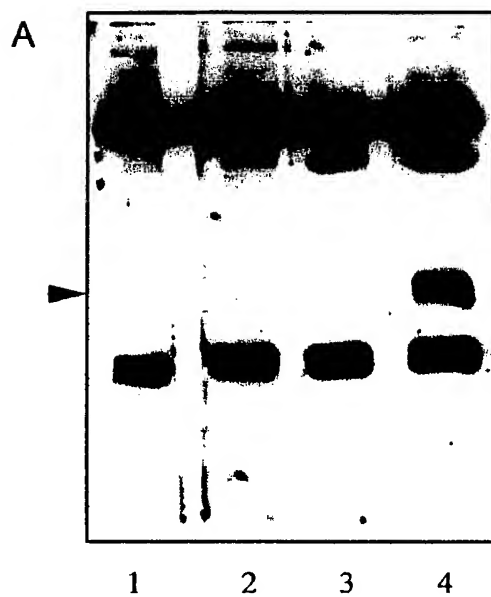
37. An immunodominant epitope derived from a *C. difficile* gene or a *C. difficile* peptide/polypeptide or a derivative or fragment or mutant or variant thereof.
- 5 38. An immunodominant epitope as claimed in claim 37 wherein the *C. difficile* peptide/polypeptide contains an amino acid sequence SEQ ID No.1 or SEQ ID No.2 or a derivative or fragment or mutant or variant thereof.
- 10 39. An immunodominant epitope as claimed in claim 35 wherein the *C. difficile* peptide/polypeptide contains an amino acid sequence SEQ ID No.3 or SEQ ID No.4 or SEQ ID No.5 or SEQ ID No.6 or SEQ ID No.7 or SEQ ID No.8 or SEQ ID No. 9 or SEQ ID No. 10 or a derivative or fragment or mutant or variant thereof.
- 15 40. A chimeric nucleic acid sequence derived from the 5' end of the *slpA* gene encoding the mature N-terminal moiety of SlpA from *C. difficile* which is immunogenic in humans.
- 20 41. A chimeric peptide/polypeptide wherein the amino acid sequence of the chimeric peptide/polypeptide is derived from the mature N-terminal moiety of SlpA from *C. difficile*.
- 25 42. A *C. difficile* peptide comprising SEQ ID No. 1.
43. A *C. difficile* peptide comprising SEQ ID No. 2.
44. A *C. difficile* gene comprising SEQ ID No. 3.
45. A *C. difficile* gene comprising SEQ ID No. 4.
- 30 46. A *C. difficile* gene comprising SEQ ID No. 5.

47. A *C. difficile* gene comprising SEQ ID No. 6.
48. A *C. difficile* gene comprising SEQ ID No. 7.
- 5 49. A *C. difficile* gene comprising SEQ ID No. 8.
50. A *C. difficile* gene comprising SEQ ID No. 9.
51. A *C. difficile* gene comprising SEQ ID No. 10.
- 10 52. The use of a *C. difficile* gene or a *C. difficile* peptide/polypeptide or a derivative or fragment or mutant or variant thereof which is immunogenic in humans in the preparation of a medicament for use in a method for the treatment or prophylaxis of *C. difficile* infection or *C. difficile* associated
- 15 disease in a host.
53. The use as claimed in claim 52 wherein the medicament which is prepared is a vaccine as claimed in any of claims 1 to 36.
- 20 54. A method for preparing a vaccine for prophylaxis or treatment of *C. difficile* associated disease, the method comprising;
- 25 obtaining a *C. difficile* gene or a *C. difficile* peptide/polypeptide or a derivative or fragment or mutant or variant thereof which is immunogenic in humans; and
- 30 forming a vaccine preparation comprised of said gene or peptide/polypeptide or derivative or fragment or mutant or variant, which is suitable for administration to a host and which when administered raises an immune response.

55. A method as claimed in claim 54 wherein the *C. difficile* peptide/polypeptide contains an amino acid sequence SEQ ID No.1 or SEQ ID No.2 or a derivative or fragment or mutant or variant thereof.
- 5 56. A method as claimed in claim 54 wherein the *C. difficile* gene contains an amino acid sequence SEQ ID No.3 or SEQ ID No.4 or SEQ ID No.5 or SEQ ID No.6 or SEQ ID No.7 or SEQ ID No.8 or SEQ ID No.9 or SEQ ID No.10 or a derivative or fragment or mutant or variant thereof.
- 10 57. A method for prophylaxis or treatment of *C. difficile* associated disease, the method comprising;
- 15 obtaining a *C. difficile* gene or a *C. difficile* peptide/polypeptide or a derivative or fragment or mutant or variant thereof which is immunogenic in humans;
- forming a vaccine preparation comprised of said gene or peptide/polypeptide or derivative or fragment or mutant or variant, and
- 20 administering the vaccine preparation to a host to raise an immune response.
- 25 58. Monoclonal or polyclonal antibodies or fragments thereof, to a *C. difficile* peptide/polypeptide or a derivative or fragment or mutant or variant thereof which is immunogenic in humans.
- 30 59. Monoclonal or polyclonal antibodies or fragments thereof, to *C. difficile* peptide/polypeptide or a derivative or fragment or mutant or variant thereof to which immunoreactivity is detected in individuals who have recovered from *C. difficile* infection.



60. Purified antibodies or serum obtained by immunisation of an animal with a vaccine according to any of claims 1 to 36.
- 5 61. The use of the antibodies or fragments as claimed in claims 58 and 59 in the preparation of a medicament for treatment or prophylaxis of *C. difficile* infection or *C. difficile* associated disease.
- 10 62. The use of the antibodies or serum as claimed in 60 in the preparation of a medicament for treatment or prophylaxis of *C. difficile* infection or *C. difficile* associated disease.
- 15 63. The use of the antibodies or fragments or serum as claimed in any of claims 58 to 60 for use in passive immunotherapy for established *C. difficile* infection.
64. The use of the antibodies or fragment or serum as claimed in any of claims 58 to 60 for the eradication of *C. difficile* associated disease.
- 20 65. Use of interleukin 12 as an adjuvant in *C. difficile* vaccine.
66. The use of humanised antibodies or serum for passive vaccination of an individual with *C. difficile* infection.



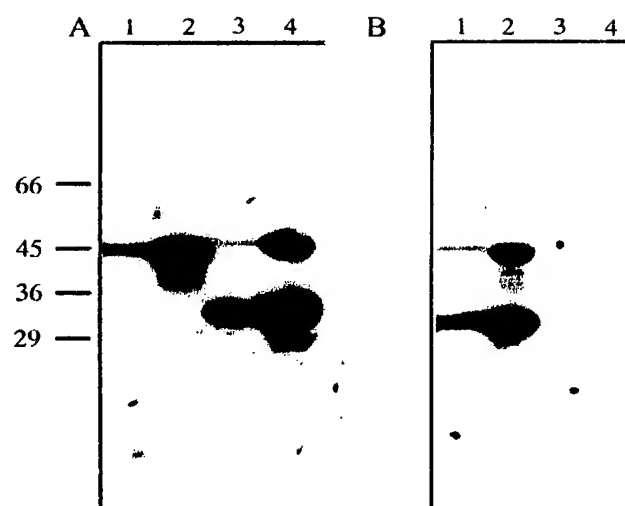


Figure 2

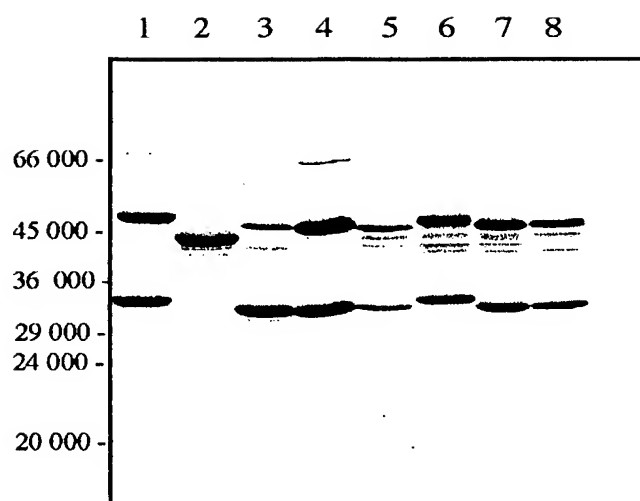


Figure 3

## SEQUENCE LISTING

<110> The Provost, Fellows & Scholars of the College of the Holy and Undivided Trinity of Queen Elizabeth, near Dublin

<120> C.difficile vaccine

<130> TRI002/C/WO

<160> 10

<170> PatentIn version 3.1

<210> 1

<211> 2157

<212> DNA

<213> Clostridium difficile

<400> 1

```
atgaataaga aaaatatagc aatagctatg tcagggtttaa cagtttttagc ttcggctgct      60
cctgttttttg ctgcaactac tggaacacaa gggttatactg tagttaaaaa cgactggaaa     120
aaagcagtaa aacaattaca ggatggacta aaagataata gtataggaaa gataactgta      180
tcttttaatg atgggggttg gggtgaagta gctcctaaaa gtgctaataa gaaagcggac     240
agagatgctg cagctgagaa gttatataat cttgttaaca ctcaattaga taaattaggt      300
gatggagatt atgttgattt ttctgtagat tataatttag aaaaaaaaaat aataactaat     360
caagcagatg cagaagcaat tgttacaaag ttaaattcac ttaatgagaa aactcttatt      420
gatatagcaa ctaaagatac ttttggaatg gttagtaaaa cacaagatag tgaaggtaaa     480
aatgttgctg caacaaaggc acttaaagtt aaagatgttg ctacatttgg tttgaagtct      540
gggtggaagcg aagatactgg atatgttggt gaaatgaaag caggagctgt agaggataag     600
tatggtaaag ttggagatag tacggcaggt attgcaataa atcttcctag tactggactt      660
gaatatgcag gtaaaggaac aacaattgat ttaataaaaa ctttaaaagt tgatgtaaca      720
gggtggttcaa cacctagtgc tgtagctgta agtgggtttg taactaaaga tgatactgat     780
ttagcaaaat caggtactat aaatgtaaga gttataaatg caaaagaaga atcaattgat      840
atagatgcaa gctcatatac atcagctgaa aatttagcta aaagatatgt atttgatcca     900
gatgaaattt ctgaagcata taaggcaata gtagcattac aaaatgatgg tatagagtct      960
aatttagttc agttagttaa tggaaaatat caagtgattt tttatccaga aggtaaaaga     1020
ttagaaacta aatcagcaaa tgatacaata gctagtcaag atacaccagc taaagtagtt     1080
ataaaagcta ataaattaaa agatttataaa gattatgtag atgatttaaa aacatataat     1140
aatacttatt caaatgttgt aacagtagca ggagaagata gaatagaaac tgctatagaa     1200
ttaagtagta aatattataa ttctgatgat aaaaatgcaa taactgataa agcagttaat     1260
```

gatatagtat tagttggatc tacatctata gttgatggtc ttgttgcatc accattagct 1320  
tcagaaaaaa cagctccatt attattaact tcaaaagata aattagattc atcagtaaaa 1380  
tctgaaataa agagagttat gaacttaaag agtgacactg gtataaatac ttctaaaaaa 1440  
gtttatttag ctgggtggagt taattctata tctaaagatg tagaaaatga attgaaaaac 1500  
atgggtctta aagttactag attatcagga gaagacagat acgaaacttc tttagcaata 1560  
gctgatgaaa taggtcttga taatgataaa gcatttgtag ttggtggtac tggattagca 1620  
gatgctatga gtatagctcc agttgcttct caacttaaag atggagatgc tactccaata 1680  
gtagttgtag atggaaaagc aaaagaaata agtgatgatg ctaagagttt cttaggaact 1740  
tctgatgttg atataatagg tggaaaaaat agcgtatcta aagagattga agagtcaata 1800  
gatagtgcaa ctggaaaaac tccagataga ataagtggag atgatagaca agcaactaat 1860  
gctgaagttt taaaagaaga tgattatttc acagatgggtg aagttgtgaa ttactttggt 1920  
gcaaaagatg gttctactaa agaagatcaa ttagtagatg ccttagcagc agcaccaata 1980  
gcaggtagat ttaaggagtc tccagctcca atcatactag ctactgatac tttatcttct 2040  
gaccaaaatg tagctgtaag taaagcagtt cctaaagatg gtggaactaa cttagttcaa 2100  
gtaggtaaag gtatagcttc ttcagttata aacaaaatga aagatttatt agatatg 2157

&lt;210&gt; 2

&lt;211&gt; 1830

&lt;212&gt; DNA

&lt;213&gt; Clostridium difficile

&lt;400&gt; 2

atgaaaaaaa gaaatttagc aatggctatg gcagctgtta ctgtagtagg ttctgctgct 60  
ccagtttttg cagcagcttc agatgtaata tctactacaag atggtacaaa tgataagtat 120  
acagtatcaa atactaaagc tagtgactta gtaaaggata ttttagcagc acaaaactta 180  
acaacagggtg cagttatttt gaacaaagat acaaaagtta ctttctatga tgcaaatgag 240  
aaagattcct caactccaac tggagataaa aaagtttatt cagaacaaac tttaactaca 300  
gctaattggaa atgaagatta tgtaaagaca acttttaaaa atttagatgc aggagaatat 360  
gctattatag atttaactta taataatgct aaaactgttg aaattaaagt agtagcagct 420  
agtgaaaaaa cagtagttgt atctagtgat gcgaaaaata gtgcaaaaga tatagctgaa 480  
aaatatgtgt ttgaagacaa agacttagaa aatgcactaa aaactataaa tgcctcagat 540  
ttcagtaaaa ctgatagtta ctatcaagta gttctttatc caaaaggaaa gagattacaa 600  
ggtttctcaa cttatagagc tacaaattat aatgaaggaa ctgcatatgg taatacacca 660  
gtaatatata ctctaaaatc tactagtaag agtaatttaa agactgcagt agaagagtta 720

```

caaaaattga atgctagtta ttctaatact acaacttttag ctggtgatga cagaatacaa      780
acagctatag agataagtaa agaataattac aataatgatg gcgagaaatc agatcattca      840
gctgatgtta aagagaatgt taaaaatggt gtattagtag gtgcaaatgc actagtagat      900
ggattagttg cggctccttt agcagcagaa aaagatgctc cactattatt aacttcaaaa      960
gataaattag attcgtcagt aaaatctgaa ataaagagag ttttagactt aaaaacttca     1020
acagaagtaa caggaaaaac agtttatata gctggtggag ttaatagtgt atctaaagaa     1080
gttgtaacag aattagaatc aatgggatta aaagttgaaa gattctcagg tgatgataga     1140
tatgaaactt ctttaaaaaat agcaggtgaa ataggcttag ataatgataa ggcttatgta     1200
gttggtggaa caggattagc agatgccatg agtatagctt cagttgcttc tactaaatta     1260
gatggtaatg gtgttgtaga tagaacaat ggacatgcta ctccaatagt tgttgtagat     1320
ggaaaagctg ataaaatatc tgatgactta gatagtttct taggaagcgc tgatgtagat     1380
ataataggtg gatttgcaag tgtatctgaa aagatggaag aagctatatc agatgctact     1440
ggtaaaggcg ttacaagagt taaaggcgac gatagacaag acactaactc tgaagttata     1500
aaaacatatt atgctaataa tactgaaata gctaaagctg cagttttaga taaagattca     1560
ggtgcttcaa gtagtgatgc aggagtattt aatttctatg tagctaaaga tggatctaca     1620
aaagaagatc aattagttga tgcattagca gtaggagctg ttgctggata taaacttgct     1680
ccagttgtat tagctactga ttctttatct tctgatcaat cggttgctat aagcaaagtt     1740
gtaggagaaa aatattctaa agatttaaca caagttggctc aaggaatagc taattcagtt     1800
ataaacaaaa tgaaagattt attagatatg                                     1830

```

```

<210> 3
<211> 2158
<212> DNA
<213> Clostridium difficile

```

```

<400> 3
atgaataaga aaaatatagc aatagctatg tcaggtttaa cagttttagc ttcggctgct      60
cctgtttttg ctgcaactac tggaacacaa gggtatactg tagttaaaaa cgactggaaa     120
aaagcagtaa aacaattaca agatggacta aaagataata gtataggaaa gataactgta     180
tcttttaatg atggggttgt ggggtgaagta gtccttaaaa gtgctaataa gaaagcggac     240
agagatgctg cagctgagaa gttatataat cttgttaaca ctcaattaga taaattaggt     300
gatggagatt atgttgattt ttctgtagat tataatttag aaaacaaaat aataactaat     360
caagcagatg cagaagcaat tgttacaaag ttaaatcac ttaatgagaa aactcttatt     420
gatatagcaa ctaaagatac ttttggaatg gttagtaaaa cacaagatag tgaaggtaaa     480

```

```

aatgttgctg caacaaaggc acttaaagtt aaagatgttg ctacatttgg tttgaagtct 540
ggtggaagcg aagatactgg atatgttggt gaaatgaaag caggagctgt agaggataag 600
tatggtaaag ttggagatag tacggcaggt attgcaataa atcttcctag tactggactt 660
gaatatgcag gtaaaggaac aacaattgat tttataaaaa ctttaaaagt tgatgtaaca 720
ggtgggttcaa cacctagtgc tgtagctgta agtgggttttg taactaaaga tgatactgat 780
ttagcaaaat caggtactat aaatgtaaga gttataaatg caaaagaaga atcaattgat 840
atagatgcaa gctcatatac atcagctgaa aatttagcta aaagatatgt atttgatcca 900
gatgaaattt ctgaagcata taaggcaata gtagcattac aaaatgatgg tatagagtct 960
aacttagttc agttagttaa tggaaaatat caagtgattt tttatccaga aggtaaaaga 1020
ttagaaacta aatcagcaaa tgatacaata gctagtcaag atacaccagc taaagtagtt 1080
ataaaagcta ataaattaaa agatttaaaa gattatgtag atgatttaaa aacatataat 1140
aatacttatt caaatgttgt aacagtagca ggagaagata gaatagaaac tgctatagaa 1200
ttaagtagta aatattataa ttctgatgat aaaaatgcaa taactgataa agcagttaat 1260
gatatagtat tagttggatc tacatctata gttgatggtc ttgttgcac accattagct 1320
tcagaaaaaa cagctccatt attattaact tcaaaagata aattagattc atcagtaaaa 1380
tctgaaataa agagagttat gaacttaaag agtgacactg gtataaatac ttctaaaaaa 1440
gtttatttag ctggtggagt taattctata tctaaagatg tagaaaatga attgaaaaac 1500
atgggtctta aagttactag attatcagga gaagacagat acgaaacttc ttagcaata 1560
gctgatgaaa taggtcttga taatgataaa gcatttgtag ttggtggtac tggattagca 1620
gatgctatga gtatagctcc agttgcttct caacttaaag atggagatgc tactccaata 1680
gtagttgtag atggaaaagc aaaagaaata agtgatgatg ctaagagttt cttaggaact 1740
tctgatgttg atataatagg tggaaaaaat agcgtatcta aagagattga agagtcaata 1800
gatagtgcaa ctggaaaaac tccagataga ataagtggag atgatagaca agcaactaat 1860
gctgaagttt taaaagaaga tgattatttc acagatgggtg aagttgtgaa ttactttgtt 1920
gcaaaagatg gttctactaa agaagatcaa ttagtagatg ccttagcagc agcaccaata 1980
gcaggtagat ttaaggagtc tccagctcca atcatactag ctactgatac tttatcttct 2040
gaccaaaatg tagctgtaag taaagcagtt cctaaagatg gtggaactaa cttagttcaa 2100
gtaggtaaag gtatagcttc ttcagttata aacaaaatga aagatttatt agatatgg 2158

```

```

<210> 4
<211> 2271
<212> DNA
<213> Clostridium difficile

```



```

<400> 4
atgaataaga aaaatatagc aatagctatg tcagggtttaa cagtttttagc ttcggctgca      60
cctgtatttg cagatgatac aaaagttgaa actggtgatac aaggatatac agtggtacaa      120
agcaagtata agaaagctgt tgaacaatta caaaaaggaa tattagatgg aagtataaca      180
gaaattaaag ttttctttga gggaacttta gcatctacta taaaagtagg ttctgagctt      240
aatgcagcag atgcaagtaa attattgttt acacaagtag ataataaact agataattta      300
ggatgatggag attatgtaga tttcttaata acttctccag gtcaagggga taaaataaact      360
acaagtaaac ttgttgcatc gaaagattta acagggtgctt cagcagatgc tataattgct      420
ggaacatctt cagcagatgg tgttggttaca aatactggag ctgctagtgg ttctactgag      480
acaaattcag caggaacaaa acttgcaatg tcagctatctt ttgacacagc atatacagat      540
tcactctgaaa ctgctggttaa gattactata aaagcagata tgaatgatac taaatttggt      600
aaagcaggtg agacaactta ttcaactggg cttacatttg aagatgggtc tacagaaaaa      660
attgttaaat taggggacag tgatattata gatataacta aagctcttaa acttactgtt      720
gttcctggaa gtaaagcaac tgttaagttt gctgaaaaaa caccaagtgc cagtgttcaa      780
ccagtaataa caaagcttag aataataaat gctaaagaag aaacaataga tattgacgct      840
agttctagta aaacagcaca agatttagct aaaaaatatg tattaataa aactgattta      900
aatactcttt ataaagtatt aaatggagat gaagcagata ctaatggatt aatagaagaa      960
gttagtgga aatatcaagt agttctttat ccagaaggaa aaagagttac aactaagagt     1020
gctgcaaagg cttcaattgc tgatgaaaat tcaccagtta aattaactct taagtcagat     1080
aagaagaaag acttaaaaga ttatgtggat gatttaagaa catataataa tggatattca     1140
aatgctatag aagtagcagg agaagataga atagaaactg caatagcatt aagtcaaaaa     1200
tattataact ctgatgatga aaatgctata tttagagatt cagttgataa tgtagtattg     1260
gttgaggaa atgcaatagt tgatggactt gtagcttctc ctttagcttc tgaaaagaaa     1320
gctcctttat tattaacttc aaaagataaa ttagattcaa gcgtaaaagc tgaaataaag     1380
agagttatga atataaagag tacaacaggt ataaataactt caaagaaagt ttatttagct     1440
ggatggagtta attctatatc taaagaagta gaaaatgaat taaaagatat gggaacttaa     1500
gttacaagat tagcaggaga tgatagatat gaaacttctc taaaaatagc tgatgaagta     1560
ggctcttgata atgataaagc atttgtagtt ggaggaacag gattagcaga tgccatgagt     1620
atagctccag ttgcatctca attaagaaat gctaattggtt aaatggattt agctgatggg     1680
gatgctacac caatagtagt tgtagatgga aaagctaaaa ctataaatga tgatgtaaaa     1740
gatttcttag atgattcaca agttgatata atagggtggag aaaacagtgt atctaaagat     1800

```

```

gttgaaaatg caatagatga tgctacaggt aaatctccag atagatatag tggagatgat 1860
agacaagcaa ctaatgcaaa agttataaaa gaatcttctt attatcaaga taacttaaat 1920
aatgataaaa aagtagttaa tttctttgta gctaaagatg gttctactaa agaagatcaa 1980
ttagttgatg ctttagcagc agctccagtt gcagcaaact ttgggtgtaac tcttaattct 2040
gatggtaagc cagtagataa agatggtaaa gtattaactg gttctgataa tgataaaaat 2100
aaattagtat ctccagcacc tatagtatta gctactgatt ctttatcttc agatcaaagt 2160
gtatctataa gttaaagtct tgataaagat aatggagaaa acttagttca agttggtaaa 2220
ggtatagcta cttcagttat aaacaaaatg aaagatttat tagatatgta a 2271

```

```

<210> 5
<211> 2158
<212> DNA
<213> Clostridium difficile

```

```

<400> 5
atgaataaga aaaatatagc aatagctatg tcagggttaa cagttttagc ttcggctgct 60
cctgtttttg ctgcaactac tggaacacaa gggtatactg tagttaaaaa cgactggaaa 120
aaagcagtaa aacaattaca agatggacta aaagataata gtataggaaa gataactgta 180
tcttttaatg atgggggttg ggggtgaagta gctcctaaaa gtgctaataa gaaagcggac 240
agagatgctg cagctgagaa gttatataat cttgttaaca ctcaattaga taaattaggt 300
gatggagatt atggtgattt ttctgtagat tataatttag aaaacaaaat aataactaat 360
caagcagatg cagaagcaat tgttacaaag ttaaattcac ttaatgagaa aactcttatt 420
gatatagcaa ctaaagatac ttttggaatg gttagtaaaa cacaagatag tggaggtaaa 480
aatgttgctg caacaaaggc acttaaagtt aaagatggtg ctacatttgg tttgaagtct 540
ggtggaagcg aagatactgg atatgttggt gaaatgaaag caggagctgt agaggataag 600
tatggtaaag ttggagatag tacggcaggt attgcaataa atcttcctag tactggactt 660
gaatatgcag gtaaaggaac aacaattgat tttaataaaa ctttaaaagt tgatgtaaca 720
ggtggttcaa cacctagtgc tgtagctgta agtggttttg taactaaaga tgatactgat 780
ttagcaaaat caggtactat aaatgtaaga gttataaatg caaaagaaga atcaattgat 840
atagatgcaa gctcatatac atcagctgaa aatttagcta aaagatatgt atttgatcca 900
gatgaaattt ctgaagcata taaggcaata gtagcattac aaaatgatgg tatagagtct 960
aatttagttc agttagttaa tggaataat caagtgattt tttatccaga aggtaaaaga 1020
ttagaaacta aatcagcaaa tgatacaata gctagtcaag atacaccagc taaagtagtt 1080
ataaaagcta ataaattaaa agatttaaaa gattatgtag atgatttaaa aacatataat 1140

```

```

aataacttatt caaatgttgt aacagtagca ggagaagata gaatagaaac tgctatagaa 1200
ttaagtagta aatattataa ttctgatgat aaaaatgcaa taactgataa agcagttaat 1260
gatatagtat tagttggatc tacatctata gttgatggtc ttgttgcatc accattagct 1320
tcagaaaaaa cagctccatt attattagct tcaaaagata aattagattc atcagtaaaa 1380
tctgaaataa agagagttat gaacttaaaag agtgacactg gtataaatac ttctaaaaaa 1440
gtttatttag ctggtggagt taattctata tctaaagatg tagaaaatga attgaaaaac 1500
atgggtctta aagttactag attatcagga gaagacagat acgaaacttc tttagcaata 1560
gctgatgaaa taggtcttga taatgataaa gcattttag taggtggtac tggattagca 1620
gatgctatga gtatagctcc agttgcttct caacttaaaag atggagatgc tactccaata 1680
gtagttgtag atggaaaagc aaaagaaata agtgatgatg ctaagagttt cttaggaact 1740
tctgatgttg atataatagg tggaaaaaat agcgtatcta aagagattga agagtcaata 1800
gatagtgcaa ctggaaaaac tccagataga ataagtggag atgatagaca agcaactaat 1860
gctgaagttt taaaagaaga tgattatttc acagatgggtg aagttgtgaa ttactttgtt 1920
gcaaaagatg gttctactaa agaagatcaa ttagtagatg ccttagcagc agcaccaata 1980
gcaggtagat ttaaggagtc tccagctcca atcatactag ctactgatac tttatcttct 2040
gaccaaaatg tagctgtaag taaagcagtt cctaaagatg gtggaactaa cttagttcaa 2100
gtaggtaaag gtatagcttc ttcagttata aacaaaatga aagatttatt agatatgg 2158

```

```

<210> 6
<211> 2217
<212> DNA
<213> Clostridium difficile

```

```

<400> 6
atgaataaga aggatatagc aatagctatg tcaggattaa cagtattagc ttctgcagca 60
cctgtatttg ctgctagtag ttttacagca gattataatt atactgtagt gcaaggaaaa 120
tatcaaaaag ttataactgg attacaagat ggtttaaaaa atggaaaaat aacaaatatt 180
gatgtaatat ttgatggaag ttcaattggt gaggtagtagc caggttctga tgctgcagct 240
gcagctacta aattaataag tttagttgat gataagttag ataacttagg tgatggaaaa 300
tacgttcaat ttaatgttac ttatactact aaatctataa taactaaagc agaattaaaa 360
aattattata atcaattaga aagtagtaaa gatagaatac ttataggaaa tgaacctcaa 420
gatacaggaa ctaaagggtct tataaaagct gatactgatg gtactactgc tgttgcagca 480
gctgcaccat tgaaattatc agatatattt acgttttagtt atgatgaagt aacagggtga 540
cttaaagcag aaccaacaag taaagtaagc gctggtaaag ttcaagggtct aaaatatgga 600

```

```

aatacaggag caactaacta tacttctgga gctgaaatat ctgttcctac tacaggctta 660
acattaactg ctgatacaac tgcaacaaca gatgtaaata tttctgatgt tatgagtgca 720
tttaaattta atggtactga tacgattagt ggattcccag ctggttcatc agcttctact 780
cttagagcaa gtataaaagt aataaatgca aaagaagaat ctatagatgt tgattcaagt 840
tcacatagaa cagctgaaga tttagctgaa aaatatgtat ttaaaccaga agatgtgaat 900
aaaacttatg aggcactgac tgatttatat aaagaaggta taacaagtaa tcttatcact 960
caagatggtg gaaaatatca agttgtttta tttgctcaag gaaagagatt aactactaaa 1020
ggagcaactg gaactttagc agatgaaaat tctcctctta aagtaacaat aaaagcagat 1080
aaagtaaaag acttaaaaga ttatgttgaa gatttaaaaa atgctaaca tggatattca 1140
aattctgttg ttgtagcagg tgaagataga atagaaacag caatagagtt aagtagcaaa 1200
tactataact ctgatgatga caatgcaata actaaagatc cagttaacaa tgttgtttta 1260
gttggttctc aagctgtagt tgatgggctt gtagcttcac ctttagcatc tgaaaaaaga 1320
gctcctttac tattaacttc agcaggaaaa ttagattcaa gtgttaaagc tgagttgaaa 1380
agagtaatgg atttaaaatc tacaacaggt gtaaatactt ctaaaaaagt ttacttagct 1440
ggtggagtaa actctatatc taaagatgta gaaaatgaat taaaagatat gggacttaaa 1500
gttacaagat tatcaggaga tgatagatat gaaacttctt tagctatagc tgatgaaata 1560
ggtcttgata atgataaagc ttttgtagtt ggaggaacag gattagcgga tgctatgagt 1620
atagctccag ttgcttctca attaagaaac tcaaattggag aacttgactt aaaaggtgat 1680
gcaactccaa tagtagttgt tgatggaaaa gctaaagata taaattctga agtaaaagat 1740
ttcttagatg attcacaagt tgatataata ggtggtgtaa atagtgtttc taaagaagta 1800
atggaagcaa tagatgatgc tactggaaaa tcacctgaga gatatagtgg agaagataga 1860
caagcaacaa atgctaaagt tataaaagaa gatgatttct ttaaaaatgg agaagttaca 1920
aacttctttg tagctaaaga tggttcaact aaagaagatc aattagtaga tgctttagca 1980
ggtgctgcaa ttgctggtaa ctttggtgta acagtagata atgaaggaaa acctacagtt 2040
gctgataaaa aagcttctcc agcaccaatt gttttagcaa cagattcttt atcttctgat 2100
caaaatgtag ctataagtaa agctgtaaat gatgacgcta atactaagaa tctagttcaa 2160
gttggttaaag gtatagctac ttcagttgta agtaaaataa aagatttatt agatatg 2217

```

&lt;210&gt; 7

&lt;211&gt; 2145

&lt;212&gt; DNA

&lt;213&gt; Clostridium difficile

<400> 7  
atgaataaga aaaacttagc aatggctatg gcagcagtta ctgttggtggg ttctgcagcg 60  
ccaatatattg cagatagtagc tacgccaggt tatactgtag tgaaaaatga ttggaaaaaa 120  
gcagtaaaac aattacaaga tgggttgaaa aataaaacta tatcaacaat aaaggtgtct 180  
tttaatggaa actctgttg agaagttaca ccagccagtt ctggagcaaa aaaagcagat 240  
agagatgctg cagctgaaaa gttatataat ttagtaaata cacaattaga taaactaggt 300  
gatggagatt acgttgactt tgaagtaact tataatttag ctactcaaat aattacaaaa 360  
gcagaagcag aggcagttct taaaaatta caacaatata atgataaagt acttataaat 420  
tctgcaacag atacagtaaa aggtatggta tctgatacac aagttgatag caaaaatgtt 480  
gcagctaacc cacttaaagt tagtgatatg tatacaatac catctgctat tactggaagt 540  
gatgattctg ggtatagtag tgctaaacca acagaaaaga ctacaagttt attgtatggt 600  
acggttggtg atgcaactgc aggtaaagca ataacagtag atacagcttc aaatgaagct 660  
tttgctggaa atggaaaggt tattgactac aataaatcat tcaaagcaac tgtacaagga 720  
gatggaacag ttaagacaag cgggggttgta cttaaagatg caagtgatat ggctgcaaca 780  
ggtactataa aagttagagt tacaagtgc aaagaagaat ctattgatgt ggattcaagt 840  
tcatatatta gtgctgaaaa ttagctaaa aaatatgtat ttaatcctaa agaggtttct 900  
gaagcttata atgcaatagt tgcattacaa aatgatggaa tagaatctga tttagtacaa 960  
ttagttaatg gaaaatatca agttattttc tatccagaag gaaaagatt agaaactaaa 1020  
tctgcagata taatagctga tgcagatagt ccagctaaaa taactataaa agctaataaa 1080  
ttaaaagatt taaaagatta tgtagatgat ttaaaaacat acaataatac ttactcaaat 1140  
gttgtaacag tagcaggaga agatagaata gaaactgcta tagaattaag tagtaaatat 1200  
tataattctg atgataaaaa tgcaataact gatgatgcag ttaataatat agtattagtt 1260  
ggatctacat ctatagttga tgggtctgtt gcatcaccat tagcttcaga aaaaacagct 1320  
ccattattat taacttcaaa agataaatta gattcatcag taaaatctga gataaaaaga 1380  
gttatgaact taaagagtga tactggtata aatacttcta aaaaagttaa tttagctggg 1440  
ggagttaatt ctatatctaa agatgtagaa gatgaattga aaaatatggg ccttaaagtt 1500  
actagattat caggagaaga cagatacgaa acttctttag caatagctga tgaaataggt 1560  
cttgataatg ataaagcatt tgtagttggg ggtactggat tggcagatgc tatgagtata 1620  
gctccagttg cttctcaact taaagatgga gatgctactc caatagtagt tgtagatgga 1680  
aaagcaaaag aaataagtga tgatgctaag agtttcttag gaacttctga tgttgatata 1740  
ataggtggaa aaaatagcgt atctaaagag attgaagagt caatagatag tgcaactgga 1800

```

aaaactccag atagaataag tggagatgac agacaagcaa ctaatgctga agtttttaaaa 1860
gaagatgatt atttcaaaga tggatgaagtt gtgaattact ttgttgcaaa agatgggttct 1920
actaaagaag atcaattagt agatgcatta gcagcagcac caatagcagg tagatttaag 1980
gagtctccag ctccaatcat actagctact gatactttat cttctgacca aaatgtagct 2040
gtaagtaaag cagttcctaa agatggtgga actaacttag ttcaagtagg taaaggtata 2100
gcttcttcag ttataaacia aatgaaagat ttattagata tgtaa 2145

```

&lt;210&gt; 8

&lt;211&gt; 2158

&lt;212&gt; DNA

&lt;213&gt; Clostridium difficile

&lt;400&gt; 8

```

atgaataaga aaaatatagc aatagctatg tcagggttaa cagtttttagc ttcggctgct 60
cctgtttttg ctgcaactac tggaacacaa gggtatactg tagttaaaaa cgactggaaa 120
aaagcagtaa aacaattaca agatggacta aaagataata gtataggaaa gataactgta 180
tcttttaatg atgggggtgt gggatgaagta gtcctaaaaa gtgctaataa gaaagcggac 240
agagatgctg cagctgagaa gttatataat cttgttaaca ctcaattaga taaattaggt 300
gatggagatt atgttgattt ttctgtagat tataatttag aaaaaaaaaat aataactaat 360
caagcagatg cagaagcaat tggtacaaag ttaaattcac ttaatgagaa aactcttatt 420
gatatagcaa ctaaagatac ttttggaatg gttagtaaaa cacaagatag tgaaggtaaa 480
aatgttgctg caacaaaggc acttaaagtt aaagatgttg ctacatttgg tttgaagtct 540
gggtggaagcg aagatactgg atatgttatt gaaatgaaag caggagctgt agaggataag 600
tatggtaaag ttggagatag tacggcaggt attgcaataa atcttcctag tactggactt 660
gaatatgcag gtaaaggaac aacaattgat tttaataaaa ctttaaaagt tgatgtaaca 720
gggtggttcaa cacctagtgc tgtagctgta agtgggtttg taactaaaga tgatactgat 780
ttagcaaaat caggtactat aaatgtaaga gttataaatg caaaagaaga atcaattgat 840
atagatgcaa gctcatatac atcagctgaa aatttagcta aaagacatgt atttgatcca 900
gatgaaattt ctgaagcata taaggcaata gtagcattac aaaatgatgg tatagagtct 960
aatttagttc agttagttaa tggaaaatat caagtgattt tttatccaga aggtaaaaga 1020
ttagaaacta aatcagcaaa tgatacaata gctagtcaag atacaccagc taaagtagtt 1080
ataaaagcta ataaattaaa agatttaaaa gattatgtag atgatttaaa aacatataat 1140
aatacttatt caaatgttgt aacagtagca ggagaagata gaatagaaac tgctatagaa 1200
ttaagtagta aatattataa ttctgatgat aaaaatgcaa taactgataa agcagttaat 1260

```

11/11

gatatagtat tagttggatc tacatctata gttgatggtc ttgttgcac accattagct 1320  
tcagaaaaaa cagctccatt attattaact tcaaaagata aattagattc atcagtaaaa 1380  
tctgaaataa agagagttat gaacttaaag agtgacactg gtataaatac ttctaaaaaa 1440  
gtttatttag ctggtggagt taattctata tctaaagatg tagaaaatga attgaaaaac 1500  
atgggtctta aagttactag attatcagga gaagacagat acgaaacttc tttagcaata 1560  
gctgatgaaa taggtcttga taatgataaa gcattttag taggtgttac tggattagca 1620  
gatgctatga gtatagctcc agttgcttct caacttaaag atggagatgc tactccaata 1680  
gtagttgtag atggaaaagc aaaagaaata agtgatgatg ctaagagttt cttaggaact 1740  
tctgatgttg atataatagg tggaaaaaat agcgtatcta aagagattga agagtcaata 1800  
gatagtgcaa ctggaaaaac tccagataga ataagtggag atgatagaca agcaactaat 1860  
gctgaagttt taaaagaaga tgattatttc acagatgggtg aagttgtgaa ttactttgtt 1920  
gcaaaagatg gtttactaa agaagatcaa ttagtagatg ccttagcagc agcaccaata 1980  
gcaggtagat ttaaggagtc tccagctcca atcatactag ctactgatac tttatcttct 2040  
gaccaaagt tagctgtaag taaagcagtt cctaaagatg gtggaactaa cttagttcaa 2100  
gtaggtaaag gtatagcttc ttcagttata aacaaaatga aagatttatt agatatga 2158

&lt;210&gt; 9

&lt;211&gt; 20

&lt;212&gt; PRT

&lt;213&gt; Clostridium difficile

&lt;400&gt; 9

Asp Lys Thr Lys Val Glu Thr Ala Asp Gln Gly Tyr Thr Val Val Gln  
1 5 10 15

Ser Lys Tyr Lys  
20

&lt;210&gt; 10

&lt;211&gt; 20

&lt;212&gt; PRT

&lt;213&gt; Clostridium difficile

&lt;400&gt; 10

Ala Thr Thr Gly Thr Gln Gly Tyr Thr Val Val Lys Asn Asp Gly Lys  
1 5 10 15

Lys Ala Val Lys  
20

SEQ ID No. 1  
(Strain 171500)

DKTKVETADQGYTVVQSKYK

SEQ ID No. 2  
(Strain 170324)

ATTTGTQGYTVVKNDGKKAVK

SEQ ID No. 3  
(Strain 171500 DNA)

ATGAATAAGAAAAATATAGCAATAGCTATGTCAGGTTTAAACAGTTTTAGCTTCGGCTGCACCT  
GTATTTGCAGATGATACAAAAGTTGAACTGGTGATCAAGGATATACAGTGGTACAAAGCAA  
GTATAAGAAAGCTGTTGAACAATTACAAAAGGAATATTAGATGGAAGTATAACAGAAATTA  
AAGTTTTCTTTGAGGGAACTTTAGCATCTACTATAAAAGTAGGTTCTGAGCTTAATGCAGCAG  
ATGCAAGTAAATTATTGTTTACACAAGTAGATAATAAACTAGATAATTTAGGTGATGGAGATT  
ATGTAGATTTCTTAATAACTTCTCCAGGTCAAGGGGATAAAATAAACTACAAGTAACTTGTG  
CATTGAAAGATTTAAACAGGTGCTTCAGCAGATGCTATAATTGCTGGAACATCTTCAGCAGATG  
GTGTTGTTACAAATACTGGAGCTGCTAGTGGTTCTACTGAGACAAATTCAGCAGGAACAAAAAC  
TTGCAATGTCAGCTATTTTTGACACAGCATATACAGATTTCATCTGAACTGCGGTAAAGATTA  
CTATAAAAGCAGATATGAATGATACTAAATTTGGTAAAGCAGGTGAGACAACCTTATTCAACTG  
GGCTTACATTTGAAGATGGGTCTACAGAAAAAATTGTTAAATTAGGGGACAGTGATATTATAG  
ATATAACTAAAGCTCTTAACTTACTGTTGTTCTGGAAGTAAAGCAACTGTTAAGTTTGCTG  
AAAAAACACCAAGTGCCAGTGTTCAACCAGTAATAACAAAGCTTAGAATAATAAATGCTAAA  
GAAGAAACAATAGATATTGACGCTAGTTCTAGTAAAACAGCACAAAGATTTAGCTAAAAAATA  
TGTATTTAATAAACTGATTTAAATACTCTTTATAAAGTATTAAATGGAGATGAAGCAGATAC  
TAATGGATTAATAGAAGAAGTTAGTGGAAAATATCAAGTAGTTCTTTATCCAGAAGGAAAAA  
GAGTTACAACCTAAGAGTGCTGCAAAGGCTTCAATTGCTGATGAAAATTCACCAGTTAAATTA  
CTCTTAAGTCAGATAAGAAGAAAGACTTAAAGATTATGTGGATGATTTAAGAACATATAAT  
AATGGATATTCAAATGCTATAGAAGTAGCAGGAGAAGATAGAATAGAACTGCAATAGCATT  
AAGTCAAAAAATATTATAACTCTGATGATGAAAATGCTATATTTAGAGATTCAGTTGATAATGT  
AGTATTGGTTGGAGGAAATGCAATAGTTGATGGACTTGTAGCTTCTCCTTTAGCTTCTGAAAA  
GAAAGCTCCTTTATTATTAACTTCAAAAGATAAATTAGATTCAAGCGTAAAGCTGAAATAAA  
GAGAGTTATGAATATAAAGAGTACAACAGGTATAAATACTTCAAAGAAAGTTTATTTAGCTG  
GTGGAGTTAATTCTATATCTAAAGAAGTAGAAAATGAATTTAAAGATATGGGACTTAAAGTT  
ACAAGATTAGCAGGAGATGATAGATATGAACTTCTCTAAAAATAGCTGATGAAGTAGGTCT  
TGATAATGATAAAGCATTTGTAGTTGGAGGAACAGGATTAGCAGATGCCATGAGTATAGCTCC  
AGTTGCATCTCAATTAAGAAATGCTAATGGTAAAATGGATTTAGCTGATGGTGATGCTACACC  
AATAGTAGTTGTAGATGGAAAAGCTAAACTATAAATGATGATGTAAGATTTCTTAGATG  
ATTCACAAGTTGATATAATAGGTGGAGAAAACAGTGTATCTAAAGATGTTGAAAATGCAATA  
GATGATGCTACAGGTAAATCTCCAGATAGATATAGTGGAGATGATAGACAAGCAACTAATGC  
AAAAGTTATAAAGAATCTTCTTATTATCAAGATAACTTAAATAATGATAAAAAAGTAGTTAA  
TTTCTTTGTAGCTAAAGATGGTTCTACTAAAGAAGATCAATTAGTTGATGCTTTAGCAGCAGCT  
CCAGTTGCAGCAAACCTTTGGTGTAACCTCTTAATTCTGATGGTAAGCCAGTAGATAAAGATGGT  
AAAGtATTAACCTGGTTCTGATAATGATAAAAAATAAATTAGTATCTCCAGCACCTATAGTATTAG  
CTACTGATTTCTTTATCTTCAGATCaAGTGTATCTATAAGTAaAGTTCTTGATAAAGATAATGG  
AGAAAACCTTAGTTCAAGTTGGTAAAGGTATAGCTACTTCAGTTATAAACAAAATGAAAGATTT  
ATTAGATATGTAA



SEQ ID No. 4  
(Strain 172450 DNA)

ATGAAAAAAGAAATTTAGCAATGGCTATGGCAGCTGTTACTGTAGTAGGTTCTGCTGCTCCA  
GTTTTTGCAGCAGCTTCAGATGTAATATCACTACAAGATGGTACAAATGATAAGTATACAGTA  
TCAAATACTAAAGCTAGTGACTTAGTAAAGGATATTTTAGCAGCACAAAACCTTAACAACAGGT  
GCAGTTATTTTGAACAAAGATACAAAAGTTACTTTCTATGATGCAAATGAGAAAGATTCTTCA  
ACTCCAACCTGGAGATAAAAAAGTTTATTTCAGAACAACTTTAACTACAGCTAATGGAAATGA  
AGATTATGTAAAGACAACCTTTAAAAAATTTAGATGCAGGAGAATATGCTATTATAGATTTAAC  
TTATAATAATGCTAAAACTGTTGAAATTTAAAGTAGTAGCAGCTAGTGAAAAACAGTAGTTGT  
ATCTAGTGATGCGAAAAATAGTGCAAAAGATATAGCTGAAAAATATGTGTTTGAAGACAAAG  
ACTTAGAAAACTGACTAAAACTATAAATGCCTCAGATTTTCAGTAAAACTGATAGTTACTATC  
AAGTAGTTCTTTATCCAAAAGGAAAGAGATTACAAGGTTTCTCAACTTATAGAGCTACAAATT  
ATAATGAAGGAACTGCATATGGTAATACACCAGTAATATTAACCTCTAAATCTACTAGTAAGA  
GTAATTTAAAGACTGCAGTAGAAGAGTTACAAAAATTGAATGCTAGTTATTCTAATACTACAA  
CTTTAGCTGGTGATGACAGAATACAAACAGCTATAGAGATAAGTAAAGAATATTACAATAAT  
GATGGCGAGAAATCAGATCATTTCAGCTGATGTTAAAGAGAATGTTAAAAATGTTGTATTAGTA  
GGTGCAATGCACTAGTAGATGGATTAGTTGCGGCTCCTTTAGCAGCAGAAAAAGATGCTCCA  
CTATTATTAACCTTCAAAAGATAAATTAGATTCGTCAGTAAAATCTGAAATAAAGAGAGTTTTA  
GACTTAAAAACTTCAACAGAAGTAACAGGAAAAACAGTTTATATAGCTGGTGGAGTTAATAG  
TGTATCTAAAGAAGTTGTAACAGAATTAGAATCAATGGGATTAAAAGTTGAAAGATTCTCAG  
GTGATGATAGATATGAAACTTCTTTAAAAATAGCAGGTGAAATAGGCTTAGATAATGATAAG  
GCTTATGTAGTTGGTGGAACAGGATTAGCAGATGCCATGAGTATAGCTTCAGTTGCTTCTACT  
AAATTAGATGGTAATGGTGTGTAGATAGAACAAATGGACATGCTACTCCAATAGTTGTTGTA  
GATGGAAAAGCTGATAAAATATCTGATGACTTAGATAGTTTCTTAGGAAGCGCTGATGTAGAT  
ATAATAGGTGGATTTGCAAGTGTATCTGAAAAGATGGAAGAAGCTATATCAGATGCTACTGGT  
AAAGGCGTTACAAGAGTTAAAGGCGACGATAGACAAGACACTAACTCTGAAGTTATAAAAAAC  
ATATTATGCTAATGATACTGAAATAGCTAAAGCTGCAGTTTTAGATAAAGATTTCAGGTGCTTC  
AAGTAGTGATGCAGGAGTATTTAATTTCTATGTAGCTAAAGATGGATCTACAAAAGAAGATCA  
ATTAGTTGATGCATTAGCAGTAGGAGCTGTTGCTGGATATAAACTTGCTCCAGTTGTATTAGCT  
ACTGATTCCTTATCTTCTGATCAATCGGTTGCTATAAGCAAAGTTGTAGGAGAAAAATATTCTA  
AAGATTTAACACAAGTTGGTCAAGGAATAGCTAATTCAGTTATAAACAAAATGAAAGATTTAT  
TAGATATG

SEQ ID No. 5  
(Strain 170324 DNA)

ATGAATAAGAAAAATATAGCAATAGCTATGTCAGGTTTAAACAGTTTTAGCTTCGGCTGCTCCT  
GTTTTTGCTGCAACTACTGGAACACAAGGTTTACTGTAGTTAAAAACGACTGGAAAAAAGCA  
GTAAAACAATTACAAGATGGACTAAAAGATAATAGTATAGGAAAGATAACTGTATCTTTTAA  
TGATGGGGTTGTGGGTGAAGTAGCTCCTAAAAGTGCTAATAAGAAAGCGGACAGAGATGCTG  
CAGCTGAGAAGTTATATAATCTTGTTAACTCAATTAGATAAATTAGGTGATGGAGATTATG  
TTGATTTTCTGTAGATTATAATTTAGAAAACAAAATAATAACTAATCAAGCAGATGCAGAAG  
CAATTGTTACAAAGTTAAATTCACTTAATGAGAAAACTCTTATTGATATAGCAACTAAAGATA  
CTTTTGGAATGGTTAGTAAACACAAGATAGTGAAGGTAAAAATGTTGCTGCAACAAAGGCA  
CTTAAAGTTAAAGATGTTGCTACATTTGGTTTGAAGTCTGGTGGAAGCGAAGATACTGGATAT  
GTTGTTGAAATGAAAGCAGGAGCTGTAGAGGATAAGTATGGTAAAGTTGGAGATAGTACGGC  
AGGTATTGCAATAAATCTTCCTAGTACTGGACTTGAATATGCAGGTAAAGGAACAACAATTGA  
TTTTAATAAACTTTAAAAGTTGATGTAACAGGTGGTTCAACACCTAGTGCTGTAGCTGTAAG  
TGGTTTGTAACTAAAGATGATACTGATTTAGCAAAATCAGGTACTATAAATGTAAGAGTTAT  
AAATGCAAAAAGAAGAATCAATTGATATAGATGCAAGCTCATATACATCAGCTGAAAATTTAG  
CTAAAAGATATGTATTTGATCCAGATGAAATTTCTGAAGCATATAAGGCAATAGTAGCATTAC  
AAAATGATGGTATAGAGTCTAACTTAGTTAGTTAGTTAATGGAAAATATCAAGTGATTTTTT  
ATCCGAAAGGTAAAGATTAGAAAACATAATCAGCAAAATGATACAATAGCTAGTCAAGATACA  
CCAGCTAAAGTAGTTATAAAAGCTAATAAATTTAAAAGATTTAAAAGATTATGTAGATGATTTA  
AAAACATATAATAATACTTATTCAAATGTTGTAACAGTAGCAGGAGAAGATAGAATAGAAAC

TGCTATAGAATTAAGTAGTAAATATTATAATTCTGATGATAAAAAATGCAATAACTGATAAAGC  
AGTTAATGATATAGTATTAGTTGGATCTACATCTATAGTTGATGGTCTTGTTCATCACCATTA  
GCTTCAGAAAAAACAGCTCCATTATTATTAACCTCAAAAAGATAAATTAGATTCATCAGTAAAA  
TCTGAAATAAAGAGAGTTATGAACTTAAAGAGTGACACTGGTATAAATACTTCTAAAAAAGTT  
TATTTAGCTGGTGGAGTTAATCTATATCTAAAGATGTAGAAAATGAATTGAAAAACATGGGT  
CTTAAAGTTACTAGATTATCAGGAGAAGACAGATACGAACTTCTTTAGCAATAGCTGATGAA  
ATAGGTCTTGATAATGATAAAGCATTGTAGTTGGTGGTACTGGATTAGCAGATGCTATGAGT  
ATAGCTCCAGTTGCTTCTCAACTTAAAGATGGAGATGCTACTCCAATAGTAGTTGTAGATGGA  
AAAGCAAAAAGAAATAAGTGATGATGCTAAGAGTTTCTTAGGAACTTCTGATGTTGATATAATA  
GGTGGAAAAAATAGCGTATCTAAAGAGATTGAAGAGTCAATAGATAGTGCAACTGGAAAAAC  
TCCAGATAGAATAAGTGGAGATGATAGACAAGCAACTAATGCTGAAGTTTTAAAGAAGATG  
ATTATTTACAGATGGTGAAGTTGTGAATTACTTTGTTGCAAAAAGATGGTTCTACTAAAGAAG  
ATCAATTAGTAGATGCCTTAGCAGCAGCACCAATAGCAGGTAGATTTAAGGAGTCTCCAGCTC  
CAATCATACTAGCTACTGATACTTTATCTTCTGACCAAAATGTAGCTGTAAGTAAAGCAGTTC  
CTAAAGATGGTGGAACCTAAGTTCAAGTAGGTAAAGGTATAGCTTCTTCAGTTATAAACA  
AAATGAAAGATTTATTAGATATGG

SEQ ID No. 6

(Strain 171448 DNA)

ATGAATAAGAAAAATATAGCAATAGCTATGTCAGGTTTAAACAGTTTTAGCTTCGGCTGCTCCT  
GTTTTTGCTGCAACTACTGGAACACAAGGTTATACTGTAGTTAAAAACGACTGGAAAAAAGCA  
GTAAAACAATTACAAGATGGACTAAAAGATAATAGTATAGGAAAGATAACTGTATCTTTTAA  
TGATGGGGTTGTGGGTGAAGTAGCTCCTAAAAGTGCTAATAAGAAAGCGGACAGAGATGCTG  
CAGCTGAGAAGTTATATAATCTTGTTAACTCAATTAGATAAATTAGGTGATGGAGATTATG  
TTGATTTTCTGTAGATTATAATTTAGAAAACAAAATAATACTAATCAAGCAGATGCAGAAG  
CAATTGTTACAAAGTTAAATTCACCTAATGAGAAAACCTTTATTGATATAGCAACTAAAGATA  
CTTTTGGAATGGTTAGTAAAACACAAGATAGTGGAGGTAAAAATGTTGCTGCAACAAAGGCA  
CTTAAAGTTAAAGATGTTGCTACATTTGGTTTGAAGTCTGGTGGAAAGCGAAGATACTGGATAT  
GTTGTTGAAATGAAAGCAGGAGCTGTAGAGGATAAGTATGGTAAAGTTGGAGATAGTACGGC  
AGGTATTGCAATAAATCTTCCTAGTACTGGACTTGAATATGCAGGTAAAGGAACAACAATTGA  
TTTTAATAAACTTTAAAAGTTGATGTAACAGGTGGTTCAACACCTAGTGTCTGTAGCTGTAAG  
TGGTTTTGTAATAAAGATGATACTGATTGTAGCAAAATCAGGTACTATAAATGTAAGAGTTAT  
AAAATGCAAAAAGAAATCAATTGATATAGATGCAAGCTCATATACATCAGCTGAAAATTTAG  
CTAAAAGATATGATTTTGATCCAGATGAAATTTCTGAAGCATATAAGGCAATAGTAGCATTAC  
AAAATGATGGTATAGAGTCTAATTTAGTTCAGTTAGTTAATGGAAAATATCAAGTGATTTTTT  
ATCCAGAAGGTAAAAGATTAGAACTAAATCAGCAATGATACAATAGCTAGTCAAGATACA  
CCAGCTAAAGTAGTTATAAAAGCTAATAAATTAAGATTATAAGATTATGTAGATGATTTA  
AAAACATATAATAATACTTATTCAAATGTTGTAACAGTAGCAGGAGAAGATAGAATAGAAAC  
TGCTATAGAATTAAGTAGTAAATATTATAATTCTGATGATAAAAAATGCAATAACTGATAAAGC  
AGTTAATGATATAGTATTAGTTGGATCTACATCTATAGTTGATGGTCTTGTTCATCACCATTA  
GCTTCAGAAAAAACAGCTCCATTATTATTAGCTTCAAAAAGATAAATTAGATTCATCAGTAAAA  
TCTGAAATAAAGAGAGTTATGAACTTAAAGAGTGACACTGGTATAAATACTTCTAAAAAAGTT  
TATTTAGCTGGTGGAGTTAATCTATATCTAAAGATGTAGAAAATGAATTGAAAAACATGGGT  
CTTAAAGTTACTAGATTATCAGGAGAAGACAGATACGAACTTCTTTAGCAATAGCTGATGAA  
ATAGGTCTTGATAATGATAAAGCATTGTAGTTGGTGGTACTGGATTAGCAGATGCTATGAGT  
ATAGCTCCAGTTGCTTCTCAACTTAAAGATGGAGATGCTACTCCAATAGTAGTTGTAGATGGA  
AAAGCAAAAAGAAATAAGTGATGATGCTAAGAGTTTCTTAGGAACTTCTGATGTTGATATAATA  
GGTGGAAAAAATAGCGTATCTAAAGAGATTGAAGAGTCAATAGATAGTGCAACTGGAAAAAC  
TCCAGATAGAATAAGTGGAGATGATAGACAAGCAACTAATGCTGAAGTTTTAAAGAAGATG  
ATTATTTACAGATGGTGAAGTTGTGAATTACTTTGTTGCAAAAAGATGGTTCTACTAAAGAAG  
ATCAATTAGTAGATGCCTTAGCAGCAGCACCAATAGCAGGTAGATTTAAGGAGTCTCCAGCTC  
CAATCATACTAGCTACTGATACTTTATCTTCTGACCAAAATGTAGCTGTAAGTAAAGCAGTTC  
CTAAAGATGGTGGAACCTAAGTTCAAGTAGGTAAAGGTATAGCTTCTTCAGTTATAAACA  
AAATGAAAGATTTATTAGATATGG

## SEQ ID No. 7

(Strain 171862 DNA)

ATGAATAAGAAAACTTAGCAATGGCTATGGCAGCAGTTACTGTTGTGGGTTCTGCAGCGCCA  
ATATTTGCAGATAGTACTACGCCAGGTTATACTGTAGTGAAAAATGATTGGAAAAAGCAGT  
AAAACAATTACAAGATGGGTTGAAAAATAAACTATATCAACAATAAAGGTGTCTTTTAATG  
GAAACTCTGTTGGAGAAGTTACACCAGCCAGTTCTGGAGCAAAAAAAGCAGATAGAGATGCT  
GCAGCTGAAAAGTTATATAATTTAGTAAATACACAATTAGATAAACTAGGTGATGGAGATTAC  
GTTGACTTTGAAGTAACTTATAATTTAGCTACTCAAATAATTACAAAAGCAGAAGCAGAGGCA  
GTTCTTACAAAATTACAACAATATAATGATAAAGTACTTATAAATTCTGCAACAGATACAGTA  
AAAGGTATGGTATCTGATACACAAGTTGATAGCAAAAATGTTGCAGCTAACCCACTTAAAGTT  
AGTGATATGTATACAATACCATCTGCTATTACTGGAAGTGATGATTCTGGGTATAGTATTGCT  
AAACCAACAGAAAAAGACTACAaGTTTATTGTATGGTACGGTTGGTGATGCAACTGCAGGTAAA  
GCAATAACAGTAGATACAGCTTCAAATGAAGCTTTTGCTGGAAATGGAAAGGTTATTGACTAC  
AATAAATCATTCAAAGCAACTGTACAAGGAGATGGAACAGTTAAGACAAGCGGGGTTGTACT  
TAAAGATGCAAGTGATATGGCTGCAACAGGTACTATAAAAGTTAGAGTTACAAGTGCAAAAG  
AAGAATCTATTGATGTGGATTCAAGTTCATATATTAGTGCTGAAAATTTAGCTAAAAAATATG  
TATTTAATCCTAAAGAGGTTTCTGAAGCTTATAATGCAATAGTTGCATTACAAAATGATGGAA  
TAGAATCTGATTTAGTACAATTAGTTAATGGAAAAATCAAGTTATTTTCTATCCAGAAGGAA  
AAAGATTAGAACTAAATCTGCAGATATAATAGCTGATGCAGATAGTCCAGCTAAAAATAACT  
ATAAAAGCTAATAAATTTAAAAGATTAAAAGATTATGTAGATGATTTAAAAACATACAATAA  
TACTTACTCAAATGTTGTAACAGTAGCAGGAGAAGATAGAATAGAACTGCTATAGAATTAA  
GTAGTAAATATTATAATTCTGATGATAAAAATGCAATAACTGATGATGCAGTTAATAATATAG  
TATTAGTTGGATCTACATCTATAGTTGATGGTCTTGTTGCATCACCATTAGCTTCAGAAAAAAC  
AGCTCCATTATTATTAACCTCAAAGATAAATTAGATTCATCAGTAAATCTGAGATAAAAAG  
AGTTATGAACTTAAAGAGTGATACTGGTATAAATACTTCTAAAAAAGTTTATTTAGCTGGTGG  
AGTTAATTCTATATCTAAAGATGTAGAAGATGAATTGAAAAATATGGGCCTTAAAGTTACTAG  
ATTATCAGGAGAAGACAGATACGAACTTCTTTAGCAATAGCTGATGAAAATAGGTCTTGATAA  
TGATAAAGCATTTGTAGTTGGTGGTACTGGATTGGCAGATGCTATGAGTATAGCTCCAGTTGC  
TTCTCAACTTAAAGATGGAGATGCTACTCCAATAGTAGTTGTAGATGGAAAAGCAAAAGAAA  
TAAGTGATGATGCTAAGAGTTTCTTAGGAACCTCTGATGTTGATATAATAGGTGGAAAAAATA  
GCGTATCTAAAGAGATTGAAGAGTCAATAGTAGTAGTGCAACTGGAAAACTCCAGATAGAATA  
AGTGGAGATGACAGACAGCAACTAATGCTGAAGTTTTAAAAGAAGATGATTATTTCAAAGA  
TGGTGAAGTTGTGAATTACTTTGTTGCAAAAGATGGTTCTACTAAAGAAGATCAATTAGTAGA  
TGCATTAGCAGCAGCACCAATAGCAGGTAGATTTAAGGAGTCTCCAGCTCCAATCATACTAGC  
TACTGATACTTTATCTTCTGACCAAAATGTAGCTGTAAGTAAAGCAGTTCCTAAAGATGGTGG  
AACTAACTTAGTTCAAGTAGGTAAAGGTATAGCTTCTTCAGTTATAAACAAAATGAAAGATTT  
ATTAGATATGTAA

## SEQ ID No. 8

(Strain 173644 DNA)

ATGAATAAGAAGGATATAGCAATAGCTATGTCAGGATTAACAGTATTAGCTTCTGCAGCACCT  
GTATTTGCTGCTAGTAGTTTTACAGCAGATTATAATTATACTGTAGTGCAAGGAAAAATATCAA  
AAAGTTATAACTGGATTACAAGATGGTTTAAAAAATGGAAAAATAACAAATATTGATGTAAT  
ATTTGATGGAAGTTCAATTGGTGAGGTAGTGCCAGGTTCTGATGCTGCAGCTGCAGCTACTAA  
ATTA AAAAGTTTAGTTGATGATAAGTTAGATAA ACTTAGGTGATGGAAAATACGTTCAATTTAA  
TGTTACTTATACTACTAAATCTATAATAACTAAAGCAGAATTA AAAAATTATTATAATCAATT  
AGAAAGTAGTAAAGATAGAATACTTATAGGAAATGAACCTCAAGATACAGGAACTAAAGGTC  
TTATA AAAAGCTGATACTGATGGTACTACTGCTGTTGCAGCAGCTGCACCATTGAAATTATCAG  
ATATATTTACGTTTAGTTATGATGAAGTAACAGGTGTACTTAAAGCAGAACCAACAAGTAAAG  
TAAGCGCTGGTAAAGTTCAAGGTCTAAAATATGGAAATACAGGAGCAACTAACTATACTTCTG  
GAGCTGAAATATCTGTTCCCTACTACAGGCTTAACATTAAGTCTGATACAACTGCAACAACAG  
ATGTAAATATTTCTGATGTTATGAGTGCATTTAAATTTAATGGTACTGATACGATTAGTGGATT  
CCCAGCTGGTTCATCAGCTTCTACTCTTAGAGCAAGTATAAAAGTAATAAATGCAAAAGAAGA  
ATCTATAGATGTTGATTCAAGTTCACATAGAACAGCTGAAGATTTAGCTGAAAAATATGTATT

TAAACCAGAAGATGTGAATAAACTTATGAGGCACTGACTGATTTATATAAAGAAGGTATAA  
CAAGTAATCTTATCACTCAAGATGGTGGAAAATATCAAGTTGTTTTATTTGCTCAAGGAAAAGA  
GATTAACACTAAAGGAGCAACTGGAACCTTAGCAGATGAAAATTCTCCTCTTAAAGTAACAA  
TAAAAGCAGATAAAGTAAAAGACTTAAAAGATTATGTTGAAGATTTAAAAAATGCTAACAAT  
GGATATTCAAATTCTGTTGTTGTAGCAGGTGAAGATAGAATAGAAACAGCAATAGAGTTAAG  
TAGCAAATACTATAACTCTGATGATGACAATGCAATAACTAAAGATCCAGTTAACAATGTTGT  
TTTAGTTGGTTCTCAAGCTGTAGTTGATGGGCTTGTAGCTTCACCTTTAGCATCTGAAAAAAGA  
GCTCCTTTACTATTAACCTCAGCAGGAAAATTAGATTCAAGTGTTAAAGCTGAGTTGAAAAAGA  
GTAATGGATTTAAAATCTACAACAGGTGTAAATACTTCTAAAAAAGTTTACTTAGCTGGTGGA  
GTAACTCTATATCTAAAGATGTAGAAAATGAATTTAAAGATATGGGACTTAAAGTTACAAG  
ATTATCAGGAGATGATAGATATGAAACTTCTTTAGCTATAGCTGATGAAATAGGTCTTGATAA  
TGATAAAGCTTTTTGTAGTTGGAGGAACAGGATTAGCGGATGCTATGAGTATAGCTCCAGTTGC  
TTCTCAATTAAAGAACTCAAATGGAGAACTTGACTTAAAGGTGATGCAACTCCAATAGTAGT  
TGTTGATGGAAAAGCTAAAGATATAAATTCTGAAGTAAAAGATTTCTTAGATGATTACAAAGT  
TGATATAATAGGTGGTGTAAATAGTGTTTCTAAAGAAGTAATGGAAGCAATAGATGATGCTAC  
TGGAAAATCACCTGAGAGATATAGTGGAGAAGATAGACAAGCAACAAATGCTAAAGTTATAA  
AAGAAGATGATTTCTTTAAAAATGGAGAAGTTACAACTTCTTTGTAGCTAAAGATGGTTCAA  
CTAAAGAAGATCAATTAGTAGATGCTTTAGCAGGTGCTGCAATTGCTGGTAACTTTGGTGTA  
CAGTAGATAATGAAGGAAAACCTACAGTTGCTGATAAAAAAGCTTCTCCAGCACCAATTGTTT  
TAGCAACAGATTCTTTATCTTCTGATCAAAATGTAGCTATAAGTAAAGCTGTAAATGATGACG  
CTAATACTAAGAATCTAGTTCAAGTTGGTAAAGGTATAGCTACTTCAGTTGTAAGTAAAATAA  
AAGATTTATTAGATATG

## SEQ ID No. 9

(Strain 170444 DNA)

ATGAATAAGAAAAATATAGCAATAGCTATGTCAGGTTTAAACAGTTTTAGCTTCGGCTGCTCCT  
GTTTTTGCTGCAACTACTGGAACACAAGGTTATACTGTAGTTAAAAACGACTGGAAAAAAGCA  
GTAAAACAATTACAAGATGGACTAAAAGATAATAGTATAGGAAAGATAACTGTATCTTTTAA  
TGATGGGGTTGTGGGTGAAGTAGCTCCTAAAAGTGCTAATAAGAAAGCGGACAGAGATGCTG  
CAGCTGAGAAGTTATATAATCTTGTTAACTCAATTAGATAAATTAGGTGATGGAGATTATG  
TTGATTTTTCTGTAGATTATAATTTAGAAAAAAAATAATACTAATCAAGCAGATGCAGAAG  
CAATTGTTACAAATTAATTCACCTAATGAGAAAACCTCTTATTGATATAGCAACTAAAGATA  
CTTTTGGAATGGTTAGTAAAAACACAAGATAGTGAAGGTAAAAATGTTGCTGCAACAAAGGCA  
CTTAAAGTTAAAGATGTTGCTACATTTGGTTTGAAGTCTGGTGGAAAGCGAAGATACTGGATAT  
GTTATTGAAATGAAAGCAGGAGCTGTAGAGGATAAGTATGGTAAAGTTGGAGATAGTACGGC  
AGGTATTGCAATAAATCTTCCTAGTACTGGACTTGAATATGCAGGTAAAGGAACAACAATTGA  
TTTTAATAAAAACCTTTAAAAGTTGATGTAAACAGGTGGTTCAACACCTAGTGCTGTAGCTGTAAG  
TGGTTTTGTAATAAGATGATACTGATTTAGCAAAATCAGGTACTATAAATGTAAGAGTTAT  
AAATGCAAAAAGAAGATCAATTGATATAGATGCAAGCTCATATACATCAGCTGAAAATTTAG  
CTAAAAGACATGTATTTGATCCAGATGAAATTTCTGAAGCATATAAGGCAATAGTAGCATTAC  
AAAATGATGGTATAGAGTCTAATTTAGTTCAAGTTAGTTAATGGAAAATATCAAGTGATTTTTT  
ATCCAGAAGGTAAAAGATTAGAAAACATAAATCAGCAAATGATACAATAGCTAGTCAAGATACA  
CCAGCTAAAGTAGTTATAAAAGCTAATAAATTTAAAGATTTAAAAGATTATGTAGATGATTTA  
AAAACATATAATAACTTATTCAAATGTTGTAAACAGTAGCAGGAGAAGATAGAATAGAAAC  
TGCTATAGAATTAAGTAGTAAATATTATAATTCTGATGATAAAAAATGCAATAACTGATAAAGC  
AGTTAATGATATAGTATTAGTTGGATCTACATCTATAGTTGATGGTCTTGTGTCATCACCATT  
GCTTCAGAAAAAACAGCTCCATTATTATTAACCTCAAAGATAAATTAGATTCATCAGTAAAA  
TCTGAAATAAAGAGAGTTATGAACTTAAAGAGTGACACTGGTATAAATACTTCTAAAAAAGTT  
TATTTAGCTGGTGGAGTTAATCTATATCTAAAGATGTAGAAAATGAATTGAAAAACATGGGT  
CTTAAAGTTACTAGATTATCAGGAGAAGACAGATACGAACTTCTTTAGCAATAGCTGATGAA  
ATAGGTCCTTGATAATGATAAAGCATTTGTAGTTGGTGCTACTGGATTAGCAGATGCTATGAGT  
ATAGCTCCAGTTGCTTCTCAACTTAAAGTAGGAGATGCTACTCCAATAGTAGTTGTAGATGGA  
AAAGCAAAAGTAAAGTAGTATGCTAAGAGTTTCTTAGGAACTTCTGATGTTGATATAATA  
GGTGGAAAAAATAGCGTATCTAAAGAGATTGAAGAGTCAATAGATAGTGCAACTGGAAAAAC  
TCCAGATAGAATAAGTGGAGATGATAGACAAGCAACTAATGCTGAAGTTTTAAAGAAGATG

ATTATTTACAGATGGTGAAGTTGTGAATTACTTTGTTGCAAAAGATGGTTCTACTAAAGAAG  
ATCAATTAGTAGATGCCTTAGCAGCAGCACCAATAGCAGGTAGATTTAAGGAGTCTCCAGCTC  
CAATCATACTAGCTACTGATACTTTATCTTCTGACCAAAATGTAGCTGTAAGTAAAGCAGTTC  
CTAAAGATGGTGGAACCTAAGTTCAAGTAGGTAAAGGTATAGCTTCTTCAGTTATAAACA  
AAATGAAAGATTTATTAGATATGA

SEQ ID No. 10

(Strain 170426 DNA)

ATGAATAAGAAAAATATAGCAATAGCTATGTCAGGTTTAAACAGTTTTAGCTTCGGCTGCTCCT  
GTTTTTGCTGCAACTACTGGAACACAAGGTTATACTGTAGTTAAAAACGACTGGAAAAAAGCA  
GTAAAAACAATTACAGGATGGACTAAAAGATAATAGTATAGGAAAGATAACTGTATCTTTTAA  
TGATGGGGTTGTGGGTGAAGTAGCTCCTAAAAGTGCTAATAAGAAAGCGGACAGAGATGCTG  
CAGCTGAGAAGTTATATAATCTTGTTAACTCAATTAGATAAATTAGGTGATGGAGATTATG  
TTGATTTTTCTGTAGATTATAATTTAGAAAAAATAATAACTAATCAAGCAGATGCAGAAG  
CAATTGTTACAAAGTTAAATTCCTTAATGAGAAAACTCTTATTGATATAGCAACTAAAGATA  
CTTTTGGAAATGGTTAGTAAAAACACAAGATAGTGAAGGTAAAAATGTTGCTGCAACAAAGGCA  
CTTAAAGTTAAAGATGTTGCTACATTTGGTTTGAAGTCTGGTGGAAGCGAAGATACTGGATAT  
GTTGTTGAAATGAAAGCAGGAGCTGTAGAGGATAAGTATGGTAAAGTTGGAGATAGTACGGC  
AGGTATTGCAATAAATCTTCTAGTACTGGACTTGAATATGCAGGTAAAGGAACAACAATTGA  
TTTTAATAAACTTTAAAAGTTGATGTAACAGGTGGTTCAACACCTAGTGCTGTAGCTGTAAG  
TGGTTTTGTAATAAGATGATACTGATTTAGCAAAATCAGGTACTATAAATGTAAGAGTTAT  
AAATGCAAAAGAAGAATCAATTGATATAGATGCAAGCTCATATACATCAGCTGAAAATTTAG  
CTAAAAGATATGTATTTGATCCAGATGAAATTTCTGAAGCATATAAGGCAATAGTAGCATTAC  
AAAATGATGGTATAGAGTCTAATTTAGTTCAGTTAGTTAATGGAAAATATCAAGTGATTTTTT  
ATCCAGAAGGTAAAAGATTAGAACTAAATCAGCAAATGATACAATAGCTAGTCAAGATACA  
CCAGCTAAAGTAGTTATAAAAGCTAATAAATTTAAAAGATTTAAAAGATTATGTAGATGATTTA  
AAAACATATAATAACTTATTCAAATGTTGTAACAGTAGCAGGAGAAGATAGAATAGAAAC  
TGCTATAGAATTAAGTAGTAAATATTATAATTCTGATGATAAAAAATGCAATAACTGATAAAGC  
AGTTAATGATATAGTATTAGTTGGATCTACATCTATAGTTGATGGTCTTGTGTCATCACCATTA  
GCTTCAGAAAAACAGCTCCATTATTATTAACCTTCAAAAGATAAATTAGATTCATCAGTAAAA  
TCTGAAATAAAGAGAGTTATGAACTTAAAGAGTGACACTGGTATAAATACTTCTAAAAAAGTT  
TATTTAGCTGGTGGAGTTAATTCTATATCTAAAGATGTAGAAAATGAATTGAAAAACATGGGT  
CTTAAAGTTACTAGATTATCAGGAGAAGACAGATACGAACTTCTTTAGCAATAGCTGATGAA  
ATAGGTCTTGATAATGATAAAGCATTTGTAGTTGGTGGTACTGGATTAGCAGATGCTATGAGT  
ATAGCTCCAGTTGCTTCTCAACTTAAAGATGGAGATGCTACTCCAATAGTAGTTGTAGATGGA  
AAAGCAAAAGAAATAAGTGATGATGCTAAGAGTTTCTTAGGAACTTCTGATGTTGATATAATA  
GGTGGAAAAAATAGCGTATCTAAAGAGATTGAAGAGTCAATAGATAGTGCAACTGGAAAAAC  
TCCAGATAGAATAAGTGGAGATGATAGACAAGCAACTAATGCTGAAGTTTTAAAGAAGATG  
ATTATTTACAGATGGTGAAGTTGTGAATTACTTTGTTGCAAAAGATGGTTCTACTAAAGAAG  
ATCAATTAGTAGATGCCTTAGCAGCAGCACCAATAGCAGGTAGATTTAAGGAGTCTCCAGCTC  
CAATCATACTAGCTACTGATACTTTATCTTCTGACCAAAATGTAGCTGTAAGTAAAGCAGTTC  
CTAAAGATGGTGGAACCTAAGTTCAAGTAGGTAAAGGTATAGCTTCTTCAGTTATAAACA  
AAATGAAAGATTTATTAGATATG

(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization  
International Bureau



(43) International Publication Date  
15 August 2002 (15.08.2002)

PCT

(10) International Publication Number  
**WO 02/062379 A3**

(51) International Patent Classification<sup>7</sup>: **C12N 15/31**,  
C07K 14/33, C12N 15/62, C07K 16/12, A61K 39/08,  
31/711

(74) Agent: **O'BRIEN JOHN A AND WELDON, Michael J**;  
c/o John A. O'Brien & Associates, Third Floor, Duncairn  
House, 14 Carysfort Avenue, Blackrock, County Dublin  
(IE).

(21) International Application Number: PCT/IE02/00017

(22) International Filing Date: 11 February 2002 (11.02.2002)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:  
2001/0137 9 February 2001 (09.02.2001) IE

(81) Designated States (*national*): AE, AG, AL, AM, AT, AU,  
AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CO, CR, CU,  
CZ, DE, DK, DM, DZ, EC, EE, ES, FI, GB, GD, GE, GH,  
GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC,  
LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW,  
MX, MZ, NO, NZ, OM, PH, PL, PT, RO, RU, SD, SE, SG,  
SI, SK, SL, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ,  
VN, YU, ZA, ZM, ZW.

(84) Designated States (*regional*): ARIPO patent (GH, GM,  
KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZM, ZW),  
Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM),  
European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR,  
GB, GR, IE, IT, LU, MC, NL, PT, SE, TR), OAPI patent  
(BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR,  
NE, SN, TD, TG).

(71) Applicant (*for all designated States except US*): **THE  
PROVOST, FELLOWS AND SCHOLARS OF THE  
COLLEGE OF THE HOLY AND UNIDIVIDED  
TRINITY OF QUEEN ELIZABETH** [IE/IE]; Near  
Dublin, College Green, Dublin 2 (IE).

(72) Inventors; and

(75) Inventors/Applicants (*for US only*): **DOYLE, Rachael**  
[IE/IE]; 19 Deerpark Avenue, Castleknock, Dublin 15 (IE).  
**KELLEHER, Dermot** [IE/IE]; 30 Royal Terrace West,  
Dun Laoghaire, County Dublin (IE). **WINDLE, Henry, J.**  
[IE/IE]; 15 Cherryfield Avenue Upper, Ranelagh, Dublin  
6 (IE). **WALSH, James, Bernard** [IE/IE]; 3 Ardlui Park,  
Blackrock, County Dublin (IE). **DEIRDRE, Ni, Eidhin**  
[IE/IE]; 15 Watkins Buildings, The Coombe, Dublin 8 (IE).

Published:

— with international search report

(88) Date of publication of the international search report:  
16 January 2003

*For two-letter codes and other abbreviations, refer to the "Guid-  
ance Notes on Codes and Abbreviations" appearing at the begin-  
ning of each regular issue of the PCT Gazette.*

(54) Title: CLOSTRIDIUM DIFFICILE VACCINE

(57) Abstract: A vaccine for the treatment or prophylaxis of *C. difficile* associated disease comprises a *C. difficile* gene or a *C. difficile* peptide/polypeptide or a derivative or fragment or mutant or variant thereof which is immunogenic in humans. The gene encodes a *C. difficile* surface layer protein, SlpA or variant or homologue thereof. The peptide/polypeptide is a *C. difficile* surface layer protein, SlpA or variant or homologue thereof. The vaccine may comprise a chimeric nucleic acid sequence.



WO 02/062379 A3

## INTERNATIONAL SEARCH REPORT

International Application No

PCT/IE 02/00017

## A. CLASSIFICATION OF SUBJECT MATTER

IPC 7 C12N15/31 C07K14/33 C12N15/62 C07K16/12 A61K39/08  
A61K31/711

According to International Patent Classification (IPC) or to both national classification and IPC

## B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

IPC 7 C12N C07K A61K

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practical, search terms used)

EPO-Internal, WPI Data, PAJ, BIOSIS, CHEM ABS Data, EMBASE

## C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category *	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	WO 99 20304 A (ORAVAX INC) 29 April 1999 (1999-04-29)  page 22	1,2,5,7, 19, 23-37, 52-54, 57-64
P,X	--- CALABI EMANUELA ET AL: "Molecular characterization of the surface layer proteins from Clostridium difficile." MOLECULAR MICROBIOLOGY, vol. 40, no. 5, June 2001 (2001-06), pages 1187-1199, XP002946325 ISSN: 0950-382X Table 1: Strain 1, 33 kDa band --- -/--	1-9,11, 19-21, 23-42, 44,52-64

☒ Further documents are listed in the continuation of box C.☒ Patent family members are listed in annex.

## \* Special categories of cited documents:

- "A" document defining the general state of the art which is not considered to be of particular relevance  
 "E" earlier document but published on or after the international filing date  
 "L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)  
 "O" document referring to an oral disclosure, use, exhibition or other means  
 "P" document published prior to the international filing date but later than the priority date claimed

- "T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention  
 "X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone  
 "Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art.  
 "&" document member of the same patent family

Date of the actual completion of the international search

16 August 2002

Date of mailing of the international search report

22 10. 2002

Name and mailing address of the ISA

European Patent Office, P.B. 5818 Patentlaan 2  
 NL - 2280 HV Rijswijk  
 Tel. (+31-70) 340-2040, Tx. 31 651 epo nl,  
 Fax: (+31-70) 340-3016

Authorized officer

Mata-Vicente, M

## INTERNATIONAL SEARCH REPORT

International Application No

PCT/IE 02/00017

C.(Continuation) DOCUMENTS CONSIDERED TO BE RELEVANT		
Category *	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
P,X	KARJALAINEN TUOMO ET AL: "Molecular and genomic analysis of genes encoding surface-anchored proteins from Clostridium difficile." INFECTION AND IMMUNITY, vol. 69, no. 5, May 2001 (2001-05), pages 3442-3446, XP002946326 ISSN: 0019-9567 Associated to Acc. No: AJ291709.	1-9,11, 19-21, 23-42, 44,52-64
A	--- CERQUETTI M ET AL: "CHARACTERIZATION OF SURFACE LAYER PROTEINS FROM DIFFERENT CLOSTRIDIUM DIFFICILE CLINICAL ISOLATES" MICROBIAL PATHOGENESIS, ACADEMIC PRESS LIMITED, NEW YORK, NY, US, vol. 28, no. 6, June 2000 (2000-06), pages 363-372, XP002946324 ISSN: 0882-4010	
A	--- MASTRANTONIO P ET AL: "Identification of Clostridium difficile genes encoding surface proteins with adhesive properties." ABSTRACTS OF THE GENERAL MEETING OF THE AMERICAN SOCIETY FOR, vol. 100, 2000, page 72 XP001002649 100th General Meeting of the American Society for Microbiology; Los Angeles, California, USA; May 21-25, 2000, 2000 ISSN: 1060-2011 the whole document -----	



# INTERNATIONAL SEARCH REPORT

International application No.  
PCT/IE 02/00017

## Box I Observations where certain claims were found unsearchable (Continuation of item 1 of first sheet)

This International Search Report has not been established in respect of certain claims under Article 17(2)(a) for the following reasons:

1. ☒ Claims Nos.:  
because they relate to subject matter not required to be searched by this Authority, namely:  
**Although claims 57, 63 and 64 are directed to a method of treatment of the human/animal body, the search has been carried out and based on the alleged effects of the compound/composition.**
2. ☒ Claims Nos.:  
because they relate to parts of the International Application that do not comply with the prescribed requirements to such an extent that no meaningful International Search can be carried out, specifically:  
**see FURTHER INFORMATION sheet PCT/ISA/210**
3. ☐ Claims Nos.:  
because they are dependent claims and are not drafted in accordance with the second and third sentences of Rule 6.4(a).

## Box II Observations where unity of invention is lacking (Continuation of item 2 of first sheet)

This International Searching Authority found multiple inventions in this international application, as follows:

**see additional sheet**

1. ☐ As all required additional search fees were timely paid by the applicant, this International Search Report covers all searchable claims.
2. ☐ As all searchable claims could be searched without effort justifying an additional fee, this Authority did not invite payment of any additional fee.
3. ☐ As only some of the required additional search fees were timely paid by the applicant, this International Search Report covers only those claims for which fees were paid, specifically claims Nos.:
4. ☒ No required additional search fees were timely paid by the applicant. Consequently, this International Search Report is restricted to the invention first mentioned in the claims; it is covered by claims Nos.:  
**(9, 11, 21, 42, 44) - (1-8, 19, 20, 23-41, 52-64) - partial**

### Remark on Protest

- ☐ The additional search fees were accompanied by the applicant's protest.
- ☐ No protest accompanied the payment of additional search fees.

## FURTHER INFORMATION CONTINUED FROM PCT/ISA/ 210

This International Searching Authority found multiple (groups of) inventions in this international application, as follows:

1. Claims: (9, 11, 21, 42, 44) - complete; (1-8, 19, 20, 23-41, 52-64) - partial

Clostridium difficile S layer protein (SlpA) comprising SEQ ID NO:1 and its corresponding gene (slpA), which comprises SEQ ID NO:3; epitopes, homologs, derivatives, variants or fragments thereof. Chimeras comprising any of the previously mentioned polynucleotides/(poly)peptides. Antibodies against those (poly)peptides. Vaccines comprising any of the former and methods for prophylaxis/treatment of C. difficile-associated diseases based on the use thereof.

2. Claims: (10, 13, 14, 17, 18, 22, 43, 46, 47, 50, 51) - complete; (1-8, 19, 20, 23-41, 52-64) - partial

Idem as subject 1, but restricted to a gene comprising SEQ ID NOs: 5, 6, 9 or 10 and a polypeptide/peptide comprising SEQ ID NO:2.

3. Claims: (12, 45) - complete; (1-8, 19, 20, 23-37, 39-41, 52-54, 56-64) - partial

Idem as subject 1, but restricted to a gene comprising SEQ ID NO:4.

4. Claims: (15, 48) - complete; (1-8, 19, 20, 23-37, 39-41, 52-54, 56-64) - partial

Idem as subject 1, but restricted to a gene comprising SEQ ID NO:7.

5. Claims: (16, 49) - complete; (1-8, 19, 20, 23-37, 39-41, 52-54, 56-64) - partial

Idem as subject 1, but restricted to a gene comprising SEQ ID NO:8.

6. Claims: (65) - complete

Use of interleukin 12 as an adjuvant in a C. difficile vaccine.

FURTHER INFORMATION CONTINUED FROM PCT/ISA/ 210

7. Claims: (66) - partial

Use of humanised antibodies for passive vaccination of an individual with C. difficile infection.

8. Claims: (66) - partial

Use of serum for passive vaccination of an individual with C. difficile infection.

FURTHER INFORMATION CONTINUED FROM PCT/ISA/ 210

Continuation of Box I.2

Please notice that:

1. The translations of the ORFs contained in appendices 1-8 are not included in the sequence listing and, therefore, the one corresponding to the first invention has not been searched. In case the applicant decided to pay additional fees, he should be aware of the fact that the same will apply to the other inventions..

2. Claims 39 and 56 refer to SEQ ID NOs:3-10 as "amino acid sequences" but, actually, they are nucleotidic sequences.

3. The sequence numbering is confusing. The sequence identity numbers mentioned in the description and claims do not correspond with those of the sequence listing (example: SEQ ID NO:1 of the description is a peptide which appears under SEQ ID NO:9 of the sequence listing).

The applicant's attention is drawn to the fact that claims, or parts of claims, relating to inventions in respect of which no international search report has been established need not be the subject of an international preliminary examination (Rule 66.1(e) PCT). The applicant is advised that the EPO policy when acting as an International Preliminary Examining Authority is normally not to carry out a preliminary examination on matter which has not been searched. This is the case irrespective of whether or not the claims are amended following receipt of the search report or during any Chapter II procedure.

# INTERNATIONAL SEARCH REPORT

Information on patent family members

International Application No

PCT/IE 02/00017

Patent document cited in search report	Publication date	Patent family member(s)	Publication date
WO 9920304 A	29-04-1999	AU 1108299 A	10-05-1999
		CA 2307331 A1	29-04-1999
		EP 1024826 A1	09-08-2000
		WO 9920304 A1	29-04-1999
		US 6214341 B1	10-04-2001
		US 2001051153 A1	13-12-2001
-----			